Technical Report 1088

Enhancing the Resource Efficiency of Live-Fire Tank Gunnery Evaluation

Monte D. Smith
Raytheon Systems Company

Joseph D. Hagman
U.S. Army Research Institute

October 1998

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the U.S. Total Army Personnel Command

EDGAR M. JOHNSON
Director

Research accomplished under contract
for the Department of the Army

Raytheon Systems Company

Technical Review by

John E. Morrison, IDA
MAJ Hubert J. Grant Jr., USARNG, Chief, Small Unit
Collective Training Team, Project SIMITAR

NOTICES

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) October 1998	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) January 1998 - September 1998
4. TITLE AND SUBTITLE Enhancing the Resource Efficiency of Live-Fire Tank Gunnery Evaluation		5a. CONTRACT OR GRANT NUMBER DASW01-94-D-0011; D.O. #0014
		5b. PROGRAM ELEMENT NUMBER 63007
6. AUTHOR(S) Monte D. Smith (Raytheon) and Joseph D. Hagman (ARI)		5c. PROJECT NUMBER A792
		5d. TASK NUMBER 2125
		5e. WORK UNIT NUMBER C04
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Raytheon Systems Company Instructional Systems Directorate 10700 Parkridge Blvd, 2nd Floor Reston, VA 20191-4356		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. MONITOR ACRONYM ARI
		11. MONITOR REPORT NUMBER Technical Report 1088

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*: This investigation reports the development of a target engagement reduction methodology that supports resource-efficient, live-fire gunnery evaluation on Tank Table VIII (TTVIII), the intermediate-level tank crew gunnery certification exercise. Through a series of multiple regression analyses, it was determined that TTVIII can be reduced from its current 10 engagements to 7 engagements. Scores on these 7 engagements can be used to predict 10-engagement-based TTVIII total scores with greater than 85% predictive accuracy. For Army National Guard (ARNG) units, the 7 engagements can be selected randomly. For Active Component (AC) units, however, the predictive subset must consist of specific engagements. For the ARNG, subsets consisting of as few as two engagements can be used to identify tank crews with little chance of achieving first-run qualification (Q1), and subsets consisting of as few as four engagements can be used to identify crews with a high probability of firing Q1. Both predictions can be made with 95% accuracy. For both the ARNG and AC, short-cut scoring models allowed the prediction of 10-engagement-based TTVIII total scores, based on subsets of any size, with calculational ease. It was concluded that more resource-efficient live-fire tank gunnery evaluation is possible in both the ARNG and AC without sacrificing evaluative validity. The magnitude of resource savings to be anticipated from use of the recommended resource-efficient methods was estimated.

15. SUBJECT TERMS

Army National Guard Active Component Armor Tank Gunnery

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	53	Joseph D. Hagman 208-334-9390

Technical Report 1088

Enhancing the Resource Efficiency of Live-Fire Tank Gunnery Evaluation

Monte D. Smith
Raytheon Systems Company

Joseph D. Hagman
U.S. Army Research Institute

Reserve Component Training Research Unit
Ruth H. Phelps, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

October 1998

Army Project Number
20363007A792

Manpower and Personnel

FOREWORD

As resources tighten, the U.S. Army National Guard is continuing to search for ways to enhance the effectiveness and efficiency of its tank gunnery training program. To this end, this report describes the results of research showing that the resource efficiency of live-fire tank gunnery evaluation on Tank Table VIII (the crew certification exercise) can be enhanced by changing its content, to include fewer engagements, and its structure, to include performance "gates" to support early qualification and remedial training decision. By making these changes, the ARNG can save roughly 20-30% of the resources normally spent on Tank Table VIII without jeopardizing its purpose or intent.

This research was conducted by the U.S. Army Research Institute for the Behavioral and Social Sciences Reserve Component Training Research Unit (ARI-RCTRU), whose mission is to improve the effectiveness and efficiency of Reserve Component training through use of the latest in training and evaluation technology. This research is supported under Work Package VIII, "Reserve Component Training Strategies (TRAIN-UP)" of ARI's Science and Technology Program for Fiscal Year 1998.

The National Guard Bureau (NGB), under Project SIMITAR (Simulation in Training for Advance Readiness) sponsored this research under a continuing Memorandum of Understanding initially signed 12 June 1985. Findings have been presented to Director, Project SIMITAR; Chief, Training Division, NGB.

Zita M Simutis
ZITA M. SIMUTIS
Technical Director

ENHANCING THE RESOURCE EFFICIENCY OF LIVE-FIRE TANK GUNNERY EVALUATION

EXECUTIVE SUMMARY

Research Requirement:

Develop a target engagement reduction methodology that supports resource-efficient, live-fire gunnery evaluation on Tank Table VIII (TTVIII), the intermediate-level, crew tank gunnery certification exercise.

Procedure:

Stepwise multiple regression routines (SPSS, 1993, 1994) were used to determine if subsets of TTVIII engagements could be used to predict TTVIII total scores. The best subsets of from two to nine engagements were identified and the predictive validity specified for each.

Findings:

The findings suggest that TTVIII can be reduced from its current 10 engagements to 7 engagements. Scores on these seven engagements can be used to predict 10-engagement-based TTVIII total scores with greater than 85% predictive accuracy. For Army National Guard (ARNG) units, the seven engagements can be selected randomly. For Active Component (AC) units, however, the predictive subset must consist of specific engagements. For the ARNG, subsets consisting of as few as two engagements can be used to identify tank crews with little chance of achieving first-run qualification (Q1), and subsets consisting of as few as four engagements can be used to identify crews with a high probability of firing Q1. Both predictions can be made with 95% accuracy. For both the ARNG and AC, short-cut scoring models allow the prediction of 10-engagement-based TTVIII total scores, based on subsets of any size, using simple calculational steps.

Use of Findings:

This research shows that enhanced resource efficiency of live-fire tank gunnery evaluation is possible for both the ARNG and AC without sacrificing the validity of the evaluation process. For the ARNG, it is estimated that about 34% of current TTVIII ammunition costs could be saved by implementing an across-the-board reduction in the number of TTVIII engagements from 10 to 7, and by implementing an early qualification program wherein exceptionally proficient crews are awarded special recognition after firing only four engagements. For the AC, savings of roughly 30% could be realized from this across-the-board reduction in the number of TTVIII engagements.

ENHANCING THE RESOURCE EFFICIENCY OF LIVE-FIRE TANK GUNNERY EVALUATION

CONTENTS

LIST OF TABLES

Page

Enhancing the Resource Efficiency of Live-Fire Tank Gunnery Evaluation

Introduction

The challenge of attaining and maintaining required combat readiness levels in the face of limited training time (e.g., Eisley & Viner, 1989) and diminishing resources (e.g., McAndrews, 1997, April) has prompted the ARNG to search for more resource-efficient ways to conduct crew-served weapons training in its combat arms units. Just exactly how to train more efficiently is not always clear, but recent approaches have relied on the use of training aids, devices, simulators and simulations (TADSS). In armor units, for instance, the ARNG is using TADSS to support the training of tank gunnery (e.g., Krug & Pickell, 1996, February). This has prompted development of a Conduct-of-Fire Trainer (COFT)-based tool for predicting live-fire gunnery performance (Hagman & Smith, 1996), a strategy for using this tool in support of TADSS-based training during weekend drill periods (Hagman & Morrison, 1996), and other TADSS-based strategies designed to maximize the payoff from training resource expenditures (e.g., Shaler, 1994; U.S. Army Armor School, 1995).

Although use of TADSS is likely to enhance the resource efficiency of tank gunnery training and evaluation, it is also possible that additional efficiencies could be achieved by streamlining the structure and content of live-fire evaluation exercises (i.e., tables). The rising cost of main gun ammunition, growing restrictions on access to live-fire range/maneuver areas, and the difficulty in transporting soldiers/crews to and from these areas suggest that the benefits of more resource-efficient live-fire tank gunnery evaluation could be substantial. Thus, an answer is needed to the question of whether the number of live-fire tank gunnery engagements can be reduced without compromising the validity of the evaluation process. The present report answers this question and describes the process followed in doing so.

We selected Tank Table VIII (TTVIII) (i.e., the crew-level gunnery proficiency certification exercise) as the target of our research. This table consists of 10 engagements, selected from a possible 12, that encompass a variety of offensive and defensive combat scenarios with single and multiple stationary and moving targets (see Appendix A) (Department of the Army, 1993). Although 10 TTVIII engagements have been fired for years to assess tank gunnery proficiency, the tightening of resources now forces a look at the question of whether fewer engagements (and resources) can be used to do the same job.

To answer this question, we examined each engagement to determine its relative predictive contribution to the TTVIII total score (i.e., that based on 10 engagements). Our expectation was that some engagements would be better predictors than others, and that this would lead to the identification of specific subsets of engagements that, in turn, would lead to the most accurate predictions. These identified subsets might consist of as few as one or two engagements or as many as nine. If these subsets, regardless of their size, produce accurate predictions, then they could be used in place of the full 10-engagement scenario for qualification purposes, thereby saving time and money without sacrificing the validity of the tank gunnery evaluation process.

In summary, four objectives guided our research, to (1) develop a target engagement reduction methodology to support resource-efficient, live-fire TTVIII tank gunnery evaluation in the ARNG, (2) identify which specific TTVIII target engagement subset(s) to use for best results, (3) estimate the magnitude of resource (e.g., time, OPTEMPO, dollars, ammunition) savings that could expected from use of these subsets for crew certification purposes, and (4) determine the generalizability of our results to the Active Component (AC).

Experiment 1

Method

Participants/Data Source

To accomplish these objectives, we analyzed the first-run TTVIII tank gunnery scores of 716 armor crews contained in Project SIMITAR's (Simulation in Training for Advanced Readiness) gunnery performance database (Smith, 1998a, 1998b). These scores (both individual engagement and total scores) were collected between 1993-1997 from the ARNG's enhanced armored and mechanized infantry brigades headquartered in Idaho, Louisiana, Mississippi, North Carolina, South Carolina, and Tennessee.

Procedure

Stepwise multiple regression routines (SPSS, 1993, 1994) were used to determine if subsets of TTVIII engagements could be used to predict TTVIII total scores. The best subsets of from 2 to 9 engagements were identified and the predictive validity specified for each.

After conducting cross-validation procedures to establish the internal consistency and generalizability of the data, we began the process of identifying optimal subsets of predictors with identification of the TTVIII engagement that best predicted the table's 10-engagement-based total score. Identification was based on part-whole Pearson product-moment coefficients of correlation (r) between individual engagement and TTVIII total scores. The best individual predictor (i.e., engagement score) was then used to construct a prediction equation of the form:

Equation 1: $$Y' = B_0 + B_1(X_1)$$

where Y' is the predicted TTVIII total score, B_0 is the intercept (or theoretical TTVIII score when the predictor variable equals zero), B_1 is the empirically derived regression coefficient linking changes in the criterion variable (i.e., TTVIII total score) with changes in the predictor variable (i.e., engagement score), and X_1 is the engagement score most highly correlated with the criterion variable.

In a simple regression model of this type, the correlation between predicted (Y') and observed TTVIII scores (Y) will equal the correlation between the predictor variable (X_1,

the single engagement score with the greatest predictive power) and the criterion variable (Y, the observed TTVIII total score). The square of this value is known as the coefficient of determination (R^2), which indicates the proportion of the variance among criterion scores that can be explained by differences in the predictor variable. If the correlation between the identified engagement score and the observed total score is $r = .70$, for instance, then the coefficient of determination would be .49 ($.70^2 = .49$), which in this instance would mean that 49% of the differences in crews' TTVIII total scores could be predicted on the basis of a single engagement score.

The next step was to examine the remaining nine engagements for the one that most significantly enhanced the predictive power of the first engagement. The extent of the new predictor's incremental power depended upon the strength of its relationship with the criterion's residual scores, after the linear effect of the first predictor was removed. After all pair-wise combinations of the original predictor with each of the remaining potential predictors were tested and the best second predictor (i.e., engagement score) was identified, a new multiple regression prediction equation was developed using the combined predictive power of the two best predictors. The new prediction equation took the following form:

Equation 2: $$Y' = B_0 + B_1(X_1) + B_2(X_2)$$

where Y', B_0, B_1, and X_1, are as defined in Equation 1, B_2 is the empirically derived regression coefficient linking changes in the TTVIII criterion variable with changes in the second predictor variable, and X_2 is the second predictor variable--the one that most strongly augments the predictive power of the original predictor.

Because it contains more than one predictor, Equation 2 is an example of a multiple regression equation. It yields a coefficient of multiple correlation (R), which is a measure of correlation between the criterion variable and a weighted linear composite of two (or more) predictor variables. When the coefficient of multiple correlation is squared (R^2), it can be interpreted in the same manner as the coefficient of determination discussed above for the case of a single predictor. It becomes, in effect, a coefficient of multiple determination.

The two-predictor multiple regression prediction equation was then fitted to the data, yielding a new set of criterion residual scores. The new set of residuals represented the criterion scores after the linear effect of the first two predictors was removed. Then the remaining engagements were examined to identify the one that most significantly enhanced predictive power when it was added to the two-predictor model to form a new three-predictor model. This step produced a new prediction equation structurally similar to Equation 2 except that it contained the term, $B_3(X_3)$, which represented the third predictor and its empirically determined regression coefficient:

Equation 3: $$Y' = B_0 + B_1(X_1) + B_2(X_2) + B_3(X_3)$$

This procedure was repeated as long as additional predictor variables (i.e., engagements) significantly enhanced the predictive power of the resulting equation. The addition of predictor variables to a multiple regression prediction equation is theoretically unlimited. The general form of the equation is:

Equation 4: $$Y' = B_0 + B_1(X_1) + B_2(X_2) + B_3(X_3) + B_n(X_n)$$

where Y', B_0, B_1, X_1,B_2, X_2, B_3, and X_3 are as defined Equations 1-3 and the term $B_n(X_n)$ represents the *n*th predictor variable (X_n) and its empirically determined regression coefficient, B_n.

Selection criteria for individual predictors. We continued to add new predictors until the point when the next predictor did not significantly ($p \leq .05$) enhance predictive power. We did not know how many engagements would be necessary to reach this point. On the one hand, it was possible that each of the 10 individual engagements would contribute a proportional amount of unique variance to the prediction equation (i.e., 10%), and that none of them could be excluded without sacrificing its unique contribution. On the other hand, it was more than likely that some engagements would have more predictive power than others. If this were the case, then the bulk of predictive accuracy might be accounted for by a subset of engagements, and once this subset was constituted, the addition of more predictors would add little predictive power. If this occurred, then one or more engagements could be excluded from the recommended engagement-reduction solution. The exact number of engagements to be dropped, however, would depend to a large extent on acceptable estimation criteria based on the subset of selected engagements.

Minimum acceptable predictive accuracy. The minimum acceptable predictive accuracy depends upon standards established by individual users. Undoubtedly, some users will demand higher levels of predictive accuracy than others. Accordingly, we decided to present sufficient information to permit users to evaluate the adequacy of engagement reduction procedures under five levels of predictive accuracy: 70%, 80%, 85%, 90%, and 95%. Seventy percent predictive accuracy served as our minimum recommended level and ninety-five percent accuracy served as the ideal, with intermediate levels of 80%, 85%, and 90% available as well. Thus, we wanted a potential user of our results to be able to specify the minimum acceptable level of predictive accuracy and then select the smallest engagement subset size satisfying that criterion. Although users may select any level of predictive accuracy from 70% to 95%, our discussions principally focused on the highest (95%) level.

Results

Descriptive Data

Table 1 presents the means and standard deviations (*SD*) for TTVIII total and individual engagement scores. These means ranged from 45.4 to 78.4 with the highest found for engagement *B*1 and the lowest found for engagement A2.

The first row of the data correlation matrix, shown in Table 2, gives part-whole coefficients of correlation between the TTVIII total score and each engagement score. Other rows in the matrix present intercorrelations between pairs of engagements. Part-whole correlations ranged from .432 (B1) to .568 (B4) with a mean of .487. Intercorrelations among engagements ranged from .076 (A5, B2) to .323 (B2, B4) with a mean of .155. The relatively low intercorrelation among engagements indicates that performance on one engagement cannot be predicted on the basis of performance on another. The relatively robust part-whole correlations, in contrast, indicate that every engagement has the potential of making its own contribution to TTVIII total score predictions.

Table 1
TTVIII Descriptive Data (N = 716)

Variable	Mean	*SD*
Total	614.0	194.0
A1	48.3	41.5
A2	45.4	41.1
A3	57.2	36.3
A4	58.3	41.6
A5	65.3	40.2
B1	78.4	38.7
B2	62.8	41.4
B3	55.6	35.4
B4	65.8	40.7
B5	76.4	40.2

Table 2
TTVIII Correlation Matrix

	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5
Total	.497	.517	.453	.540	.455	.432	.497	.462	.568	.451
A1		.164	.162	.133	.205	.133	.132	.168	.171	.131
A2			.180	.191	.202	.179	.175	.138	.187	.106
A3				.189	.147	.102	.146	.141	.151	.079
A4					.265	.190	.161	.131	.210	.146
A5						.106	.076	.079	.078	.085
B1							.083	.111	.134	.077
B2								.186	.323	.150
B3									.241	.143
B4										.287

Split-Half Cross-Validation

A split-group, cross-validation design (Tatsuoka, 1969) was used to test for internal consistency and generalizability of the data to other ARNG tank crew samples. Approximately half of the 716 tank crews were assigned at random (by SPSS Version 6.1 for Windows) to the normative group and the other half were assigned to the cross-validation group. A series of least squares multiple regression prediction equations was then developed for the normative group. Stepwise procedures were used to select optimal

subsets of 1, 2, 3, 4, 5, 6, 7, 8 and 9 predictor variables with a separate equation developed for each subset size. All prediction equations were statistically significant, producing Multiple *R's* ranging from .58 (based on 1 predictor) to .98 (based on 9 predictors) and F ratios ranging from 183.57 (df = 1, 353) to 967.52 (df = 9, 345), with a rejection region of .0001 used for all equations.

The equations for the normative group were then tested on the cross-validation group and the accuracy of predictions for the two groups compared. Results revealed that, regardless of the number of predictors involved, models developed from normative group data accounted for a comparable amount of TTVIII total score variance in the cross-validation group. Tests for differences between Multiple *R's* (Hayes, 1963) produced nonsignificant z values ranging from < 1 to 1.68. Thus, the predictive models were found to be valid and, therefore, likely to maintain similar efficiency when used to predict the TTVIII total scores of other ARNG tank crew samples. Given the similar outcomes of the separate group analyses, along with our desire to obtain the best possible predictions from the largest sample size possible, subsequent analyses were conducted on pooled-group data (N = 716).

Development of Pooled-Group Prediction Equations

Using previously described stepwise multiple regression procedures, we developed prediction equations for the best subsets of 1, 2, 3, 4, 5, 6, 7, 8, and 9 engagements (see Table 3). The order of engagement entry into the equations is shown in the first column. The equations themselves are shown in Table 4.

Table 3
Stepwise Multiple Regression Results

Order of Entry	Multiple R	Adjusted R^2	SE	df	F	p
1 B4	.568	.321	159.80	1, 714	339.77	.0001
2 A4	.713	.507	136.23	2, 713	368.43	.0001
3 A1	.801	.640	116.47	3, 712	423.89	.0001
4 A2	.859	.737	99.53	4, 711	501.29	.0001
5 B3	.891	.792	88.51	5, 710	545.02	.0001
6 B2	.915	.837	78.42	6, 709	611.18	.0001
7 A5	.939	.880	67.21	7, 708	749.92	.0001
8 B1	.960	.921	54.37	8, 707	1,049.25	.0001
9 B5	.981	.962	39.03	9, 706	1,883.92	.0001

Prediction equations for every subset size were statistically significant, producing Multiple *R's* ranging from .57 (based on 1 predictor) to .98 (based on 9 predictors) and F ratios ranging from 339.77 (df = 1, 714) to 1,883.92 (df = 9, 706). The first predictor to enter the equation (B4) had the highest zero-order correlation (r = .568) with the criterion (see Table 2). This predictor alone accounted for almost one third of TTVIII total score variation (32.1%). The addition of the second predictor (A4) boosted the proportion of explained variance to 50.7%, and the proportion increased significantly with the addition of each subsequent predictor.

Table 4

Prediction Equations for Subset Sizes 1 to 9

Subset Size	Prediction Equation
1	$Y' = 435.9196 + 2.7063(B4)$
2	$Y' = 344.9851 + 2.0561(A4) + 2.2661(B4)$
3	$Y' = 288.7183 + 1.7377(A1) + 1.8807(A4) + 2.0006(B4)$
4	$Y' = 254.9244 + 1.5543(A1) + 1.5276(A2) + 1.6594(A4) + 1.7909(B4)$
5	$Y' = 209.3420 + 1.4230(A1) + 1.4454(A2) + 1.5898(A4) + 1.3429(B3) + 1.5627(B4)$
6	$Y' = 178.7824 + 1.3730(A1) + 1.3463(A2) + 1.5151(A4) + 1.0653(B2) + 1.2269(B3) + 1.2806(B4)$
7	$Y' = 135.8675 + 1.2125(A1) + 1.2038(A2) + 1.2866(A4) + 1.0670(A5) + 1.0640(B2) + 1.2145(B3) + 1.3049(B4)$
8	$Y' = 78.1757 + 1.1407(A1) + 1.0852(A2) + 1.1561(A4) + 1.0437(A5) + 1.0550(B1) + 1.0624(B2) + 1.1584(B3) + 1.2473(B4)$
9	$Y' = 32.9177 + 1.0832(A1) + 1.0670(A2) + 1.0954(A4) + 1.0186(A5) + 1.0444(B1) + 1.0223(B2) + 1.0951(B3) + 1.0234(B4) + 0.9902((B5)$

Random Subsets of Engagements

As shown in Table 3, the order in which engagements were entered into the stepwise routine was: B4 → A4 → A1 → A2 → B3 → B2 → A5 → B1 → B5 → A3. To obtain optimal predictive accuracy, the best combination of two predictors would be B4 + A4. The best combination of three predictors would be B4 + A4 + A1. For a subset of four predictors, the next engagement in the sequence (A2) would be added to the first three. In this manner, subsets of any desired size could be created.

Unfortunately, knowledge of which subsets of engagements serve as the best TTVIII total score predictors introduces the possibility of units "training to the test" in order to save time, especially if any of these subsets were eventually to take the place of the current 10-engagement TTVIII scenario. To discourage training to the test and, thereby, promote the training of the widest variety of engagements possible in preparation for TTVIII firing, engagements to be included in any particular subset could be selected at random. A random selection process would necessitate training on all relevant engagements because crews would not know beforehand which particular subset(s) of engagements would be included on TTVIII. The predictive accuracy of randomly selected subsets of engagements, however, is not known. So the question is, then, whether predictive accuracy would be seriously reduced or not if TTVIII subsets were selected at random

To answer this question, random subsets of engagements were constituted for subset sizes ranging from two to nine. This was accomplished by labeling 10 coins A1 through A5 and B1 through B5. The coins were placed in a hat and drawn (blindly) to constitute a random subset of engagements of the desired size. Once a subset was constituted, drawn coins were replaced, the coins were shaken to redistribute them physically inside the hat, and the process was repeated until a total of five random subsets were constituted for each subset size from two to nine. Subsets of size six or greater were created by

random exclusion. That is, to create a subset of size 6, four engagements were drawn randomly and excluded. The six engagements remaining in the hat became the subset. For subsets of size seven, three engagements were randomly excluded, and so on. This produced five 2-engagement random subsets, five 3-engagement random subsets, five 4-engagement random subsets, and so on, up to and including five 9-engagement random subsets. In all, 40 random subsets were constructed, 5 for each of 8 possible subset sizes.

For each of the 40 random subsets, multiple regression procedures were used to construct prediction equations. For each subset size, the predictive power of random subsets of engagements was then compared to the predictive power of the best possible combination of engagements identified statistically.

Random subsets of $N = 2$ through 6. The predictive power of randomly constituted subsets of 1, 2, 3, 4, 5, and 6 engagements was tested against the predictive power of the best subsets of predictors of each corresponding size. For each subset size, z tests between the mean Multiple R for the random subsets and the Multiple R for the best predictors indicated that the latter were superior. The z scores were 2.64, 2.89, 2.91, 2.57, and 2.11 for subsets sizes 2 through 6, respectively. The first four z values were significant at $p < .01$, and the last one was significant at $p < .05$. Details of these analyses can be found in Appendix B.

Random subsets of $N = 7$. Table 5 presents the results for subsets of $N = 7$. The first five rows present multiple regression results for the 5 random subsets. Means in the sixth line of the table are based upon the five individual random subsets. The cell under the "p" column for the "Mean" row is blank because it is meaningless to calculate a mean probability level in this situation. The last line in the table provides multiple regression results based upon the seven best predictors (B4 + A4 + A1 + A2 + B3 + B2 + A5).

Seven-predictor random subsets accounted, on average, for 85.8% of criterion (i.e., TTVIII total score) variance and produced *SEs* in the 70s, along with F ratios averaging over 600. By comparison, the 7 best predictors accounted for 88.0% of criterion variance. A test between the mean Multiple R for the random subsets and the Multiple R for the seven best predictors indicated that the latter were comparable to random subsets, $z = 1.72, p > .05$. Although the best seven engagements were numerically better predictors, their 2.2% advantage was not statistically reliable. Thus, randomly selected subsets of size $N = 7$ were as effective in predicting TTVIII total scores as the seven best predictors identified on the basis of regression routines.

Eight-predictor random subsets accounted, on average, for 91.3% of criterion variance and produced *SE*s in the 50s, along with F ratios approaching 1,000. By comparison, the eight best predictors accounted for 92.1% of criterion variance. A test between the mean Multiple R for the random subsets and the Multiple R for the eight best predictors indicated that the predictive power of the best predictors did not differ significantly from that of random subsets of the same size, $z < 1, p > .05$. The best engagements enjoyed an advantage of only 0.8% which was statistically nonsignificant. Thus, randomly selected subsets of size $N = 8$ were as effective in predicting TTVIII total

score as the eight best predictors identified on the basis of stepwise multiple regression routines.

Table 5

Random Subsets of N = 7 vs. the Seven Best Predictors

Excluded Predictors	Multiple R	Adjusted R^2	$F(7, 708)$	p	SE
A5, B4, B5	.931	.866	658.99	<.0001	71.11
B2, B3, B4	.919	.843	548.15	<.0001	76.94
A3, B2, B4	.927	.857	613.27	<.0001	73.35
A2, A5, B2	.930	.863	642.98	<.0001	71.87
A1, A5, B3	.928	.860	628.13	<.0001	72.60
Mean	.927	.858	618.30		73.17
Best 7	.939	.880	749.92	<.0001	67.21

Random subsets of N = 8. Table 6 presents the results for subsets of $N = 8$. The last line in the table provides multiple regression results based upon the eight best predictors (B4 + A4 + A1 + A2 + B3 + B2 + A5 + B1).

Table 6

Random Subsets of N = 8 vs. the Eight Best Predictors

Excluded Predictors	Multiple R	Adjusted R^2	$F(8, 707)$	p	SE
A1, A4	.954	.909	895.33	<.0001	58.47
A1, A3	.955	.911	911.89	<.0001	57.99
B1, B2	.957	.916	971.50	<.0001	56.33
A4, B3	.958	.918	995.80	<.0001	55.70
B2, B5	.956	.913	940.79	<.0001	57.16
Mean	.956	.913	943.06		57.13
Best 8	.960	.921	1,049.25	<.0001	54.37

Random subsets of N = 9. Table 7 presents the results for subsets of $N = 9$. The last line in the table provides multiple regression results based upon the nine best predictors (B4 + A4 + A1 + A2 + B3 + B2 + A5 + B1 + B5).

Table 7

Random Subsets of N = 9 vs. the Nine Best Predictors

Excluded Predictor	Multiple R	Adjusted R^2	$F(8, 707)$	p	SE
B5	.977	.953	1,621.09	<.0001	41.94
A4	.976	.952	1,571.85	<.0001	42.56
A5	.978	.956	1,731.24	<.0001	40.65
B4	.979	.958	1,806.96	<.0001	39.82
B1	.976	.953	1,595.23	<.0001	42.27
Mean	.977	.954	1,665.27		41.45
Best 9	981	.962	1,883.92	<.0001	39.03

Nine-predictor random subsets accounted, on average, for 95.4% of criterion variance and produced *SE*s in the 40s, along with *F* ratios of over 1,000. By comparison, the nine best predictors accounted for 96.2% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple R for the nine best predictors indicated that the predictive power of the best predictors did not differ significantly from that of random subsets of the same size, $z < 1, p > .05$. Thus, randomly selected subsets of size $N = 9$ were as effective in predicting TTVIII total scores as the nine best predictors identified on the basis of stepwise multiple regression routines.

For subsets consisting of from two to six engagements, the greatest predictive power is achieved by following the engagement selection strategy supported by stepwise multiple regression procedures. For larger subsets, however, randomly selected engagements have about the same predictive power as the best engagements. The practical implication of this is that crews can be trained on all 10 TTVIII engagements (plus a variety of others not included in the table) but tested on random subsets of at least seven engagements TTVIII total scores can then be predicted based upon the administered random subset. The accuracy of the resulting estimates will depend upon subset size with predictive accuracy equaling or exceeding 95%, 90%, and 85% with nine, eight, or seven randomly selected engagements, respectively.

A Shortcut Prediction Model

For subset sizes of six or smaller, the preferred course or action would be to use engagements identified by stepwise multiple regression procedures. For larger subsets, particular engagements are less important. Randomly selected subsets of engagements seem to work as well as subsets identified by stepwise procedures as long as at least seven engagements are used.

Regardless of the size of the subset, however, and regardless of whether engagements in the subset are selected randomly or statistically, the user is still saddled with a cumbersome prediction process when it comes to actually implementing the predictive model. The commander who wants to trim one engagement from the standard 10-engagement TTVIII scenario, for example, must administer nine engagements, score them, and then multiply each engagement score by its respective regression coefficient from Table 4. The resulting nine weighted scores then must be summed and added to the nine-engagements' prediction equation constant (32.92) in order to arrive at the predicted 10-engagement TTVIII total score. For nonresearchers, this could be an overwhelming requirement, especially when performed in the field. Some commanders might even argue that calculating the predicted 10-engagement TTVIII total score based on nine engagements would take more time and effort than shooting all 10 TTVIII engagements in the first place.

All of which raises the question of whether it is possible to develop a shortcut prediction model that can be easily implemented in field settings with minimal sacrifice of predictive accuracy. One approach might be to drop regression coefficients altogether. We know that as subset size approaches 10, regression coefficients become progressively

uniform, and hence unneeded. In fact, when all 10 possible predictors enter the prediction equation, coefficients approximate unity (i.e., 1.0). From the results above, we know that when subset size is six or greater, engagements are basically interchangeable, which means that regression coefficients are also interchangeable, and hence unneeded.

An examination of the regression coefficients produced in the seven-, eight-, and nine-engagement prediction models revealed little variation in their magnitudes. (see Table 4.) For the nine-engagement prediction model, for instance, coefficients hovered around 1.0 with a mean of 1.048846. If all the coefficients are essentially identical, it should be possible to eliminate them and substitute a procedure that weighs each engagement equally and eliminates the constant.

If regression coefficients could be dropped altogether without undue sacrifice of predictive precision, a possible shortcut prediction model could be reduced to three steps:

1: Add the engagement scores of the desired subset size.
2: Divide the sum by N_{sub}, the number of engagements in the subset.
3: Multiply the quotient by 10.

In this manner, each engagement is weighted equally (by dividing by N) and the mean of all engagements in the subset is extrapolated to a 10-engagement TTVIII total score (by multiplying by 10). The shortcut procedure has the effect of lumping the variance from all available engagements into a single predictor.

The efficacy of the proposed shortcut prediction model was tested by constructing a series of shortcut predictor variables. For each subset size (from $N = 2$ through 9), six shortcut predictor variables were constructed. The first shortcut variable for each subset size was based on the best set of engagements identified in the stepwise regression procedures. For example, for the N = 2 subset, the first shortcut predictor variable was calculated by this procedure:

$$[(B4 + \mathrm{A4})/2] \times 10$$

Thus, if a crew fired a score of 55 on engagement B4 and a score of 97 on engagement A4, its (shortcut) predicted TTVIII total score would be 760. This shortcut score was then used as an independent variable to predict actual TTVIII scores.

The other five shortcut predictor variables (for each subset size) were based on the randomly constituted engagement subsets described earlier. These subsets are listed in Appendix B for subset sizes two through six, and in Tables 5, 6, and 7 for subset sizes seven, eight, and nine, respectively. The first random shortcut predictor variable, for example, was created with the following procedure:

$$[(\mathrm{A3} + B5)/2] \times 10$$

Each random shortcut predictor variable was then used to predict actual TTVIII scores.

Regardless of subset size, the primary interest was in whether the shortcut method could be used to predict TTVIII total scores with the same degree of accuracy as models incorporating individual engagement scores and regression coefficients. A secondary interest was in the relative effectiveness (for each subset size) of shortcut predictions based on random subsets of engagements vs. shortcut predictions based on subsets consisting of the best possible engagements.

The results of the shortcut test appear in Table 8. The first column under the "Full Regression Models" heading shows R^2 values for each subset size for the best engagement predictors as determined by stepwise multiple regression procedures. The second column under the Full Regression Models heading shows mean R^2 values from five randomly constituted subsets of engagements. The data in the two columns under Full Regression Models are derived from Appendix B and from Tables 5-7. By comparing these two columns, it can be seen that the best predictors consistently outperform randomly selected predictors up to subset size $N = 7$, at which point random subsets do not differ statistically from the corresponding subsets consisting of the best possible predictors.

The last two data columns in Table 8 are based on shortcut regression models. The next-to-last column presents R^2 values using shortcut models based on the best predictors for each subset size. The last column contains mean R^2 value obtained from five shortcut regression models based upon randomly selected subsets of predictors. The values in the last two columns resemble those in the first two columns, with the best subsets outperforming randomly constituted subsets until subset size $N = 7$ is reached, after which point the values do not differ significantly.

Table 8
R^2 Values for Full Regression Models vs. Shortcut Regression Models

	Full Regression Models (R^2)		Shortcut Regression Models (R^2)	
Subset Size	Best Subsets	Random Subsets	Best Subsets	Random Subsets
2	.507	.403	.507	.401
3	.640	.542	.639	.543
4	.737	.660	.736	.658
5	.792	.734	.792	.732
6	.837	.802	.833	.799
7	.880	.858	.879	.856
8	.921	.913	.921	.912
9	.962	.954	.960	.954

Table 8 reveals that the shortcut prediction method can be used successfully with reduced subsets of any size. It also reinforces the earlier finding that predictions based on small subsets should use the engagements identified in the stepwise regression procedure as the best possible predictors, whereas predictions based on subsets of N = 7 or more can use engagements selected at random.

An Alternate Definition of Predictive Accuracy

To this point, we have defined predictive accuracy as R^2, the proportion of criterion variance accounted for by a weighted linear model based upon a subset of engagement scores. R^2 is obtained by squaring the correlation coefficient between observed and predicted TTVIII scores. As such, it is a measure of the goodness of fit of a linear model. R^2 has precise meaning to researchers, but it is less meaningful to most others. Fortunately, it is not the only definition of predictive accuracy.

Instead of predicting specific scores on TTVIII, it is also possible to predict crew qualification status. This prediction can have more intuitive appeal to military leaders because they are often more interested in qualification vs. nonqualification than in specific scores. The important thing to them is whether actual and predicted scores are above or below 700, the minimum cutoff score for TTVIII qualification. Efforts to predict qualification status on the basis of subsets of engagements have four possible outcomes:

1. A crew is predicted to qualify and does.
2. A crew is predicted to qualify but does not.
3. A crew is predicted not to qualify and does not.
4. A crew is predicted not to qualify but does.

Outcomes 1 and 3 are predictive successes. Outcomes 2 and 4 are predictive failures. A measure of predictive accuracy can be defined as:

$$[(1) + (3)] / [(1) + (2) + (3) + (4)]$$

This is a stringent definition of predictive accuracy because predictions of qualification vs. non-qualification are based on scores from all parts of the predictor score distribution. As an example, consider the shortcut prediction method with subset size $N = 2$. The scores from two engagements are summed, divided by 2, and multiplied by 10, producing a distribution of scores that is likely to range from 0 to 1,000, with a mean of about 614 (see Table 1). Crews that obtain a score close to 0 on this shortcut predictor are unlikely to qualify, whereas crews with scores approaching 1,000 have an excellent chance of qualifying. Thus, predictions based on extreme predictor scores are likely to produce outcomes of Types 1 and 3 and are likely to be predictive successes.

With scores that fall near the midpoint of the shortcut predictor distribution, however, predictions are more difficult. A crew with a score of 700 on the shortcut predictor, for example, could easily fall on either side of the 10-engagement-based qualification cutoff. For this reason, dichotomous criterion outcomes are most accurately predicted when they are based on extreme scores in the tails of the shortcut predictor's distribution. A practical application of this principle is that it should be possible to make directional predictions, based on subsets of engagements, with high levels of predictive accuracy.

Identifying Crews for Early Tank Table VIII Elimination

Of the 716 crews included in the SIMITAR database, 58.9% failed to qualify on their first-run (i.e., did not Q1). Armed with this knowledge, and knowing nothing more about any particular crew, the best guess that could be made regarding the outcome of a given crew's first-run qualification attempt would be to predict failure. A Q1 failure prediction would be correct about 60% of the time. Is it possible that crews with little chance of Q1 success could be identified early in the evaluation process on the basis of their performance on key predictive engagements? If so, these crews could be recalled to the starting line, thereby saving the ammunition they would have fired on subsequent engagements. Recalled crews could then be sent for device-based remedial training and allowed to return to the live-fire range only when device-based performance indicated a satisfactory probability of success (Hagman & Smith, 1996).

Formalizing early elimination predictions. For any given subset size (N_{sub}), the minimum score (E_{Elim}) necessary in order to avoid early elimination can be predicted from the general equation:

Equation 5: $$E_{Elim} = ([700 - (1.65 * SE)] / 10)* N_{sub}$$

where 700 represents the minimum 10-engagement-based TTVIII score required for qualification, 1.65 is the normal deviate (in a one-tailed directional test) for 95% confidence, SE is the standard error of estimate, and N_{sub} is the subset size (i.e., number of engagements) upon which the prediction is based, with a potential range in this instance of from two to nine. Crews failing to equal or exceed the stipulated minimum cutoff score could be eliminated from firing further TTVIII engagements with 95% confidence that their eventual score would have been less than 700 if they had continued to fire all the engagements. The *SE*s are based on stepwise regression procedures (see Table 3). Table 9 presents the minimum E_{Elim} score for each subset size. After firing the number of engagements listed in the far left column, crews failing to accumulate at least the number of points specified in the far right column would have no more than a 5% subsequent chance of Q1.

Table 9

Minimum E_{Elim} Values to Avoid Early Elimination

Subset Size	Prediction Equation	Minimum E_{Elim}
2	E_{Elim} = [700 - (1.65 * 136.2)] / 10 * 2	95
3	E_{Elim} = [700 - (1.65 * 116.5)] / 10 * 3	152
4	E_{Elim} = [700 - (1.65 * 99.5)] / 10 * 4	214
5	E_{Elim} = [700 - (1.65 * 88.5)] / 10 * 5	277
6	E_{Elim} = [700 - (1.65 * 78.4)] / 10 * 6	342
7	E_{Elim} = [700 - (1.65 * 67.2)] / 10 * 7	412
8	E_{Elim} = [700 - (1.65 * 54.4)] / 10 * 8	488[a]
9	E_{Elim} = [700 - (1.65 * 39.0)] / 10 * 9	572[b]

[a] Crews are mathematically eliminated with a score < 500.
[b] Crews are mathematically eliminated with a score < 600.

Testing the early elimination model. Scores from the SIMITAR database were used to test the early elimination prediction model and the equations in Table 9. All subsets were based on optimal predictors as identified by stepwise regression procedures (see Table 3). For example, using a subset size of two, each crew's scores on engagements B4 and A4 were summed. Summed scores were then partitioned into those < 95 and those ≥ 95. No predictions were made for crews with scores ≥ 95. Crews with scores < 95, in contrast, were predicted to have no more than a 5% chance of Q1.

This procedure was then repeated for each subset size, using appropriate cutoff scores from Table 9. For example, crews with scores of less than 152, on the basis of three engagements, were flagged as unlikely to Q1. With four engagements, the cutoff point was 214, and so on.

The accuracy of these failure-to-qualify predictions were then tested by noting whether crews scoring below the stipulated cutoff points actually failed to qualify, based on their 10-engagement-based TTVIII total score. The pertinent question was what proportion of the crews identified by this procedure actually failed to Q1. Table 10 shows the actual performance outcomes (i.e., either < 700 or ≥ 700) of crews flagged as unlikely to fire Q1.

Table 10
Accuracy of Early Elimination Predictions

Subset Size	Actual TTVIII Score < 700[a]	Actual TTVIII Score ≥ 700[a]	Predictive Accuracy
2	*N* = 200 (27.9%) Mean = 409.8	*N* = 14 (2.0%) Mean = 737.6	200/214 = 93.5%
3	*N* = 247 (34.5%) Mean = 421.0	*N* = 15 (2.1%) Mean = 734.5	247/262 = 94.3%
4	*N* = 314 (43.9%) Mean = 443.2	*N* = 25 (3.5%) Mean = 727.0	314/339 = 92.6%
5	*N* = 323 (45.1%) Mean = 443.8	*N* = 23 (3.2%) Mean = 729.5	323/346 = 96.4%
6	*N* = 323 (45.1%) Mean = 444.3	*N* = 13 (1.8%) Mean = 727.2	323/336 = 96.1%
7	*N* = 336 (46.9%) Mean = 448.1	*N* = 6 (0.8%) Mean = 721.2	336/342 = 98.2%
8	*N* = 332 (46.4%) Mean = 443.3	*N* = 0 (0.0%) Mean = na	332/332 = 100.0%
9	*N* = 346 (48.3%) Mean = 449.7	*N* = 0 (0.0%) Mean = na	346/346 = 100.0%

[a] Percentages in these column are based on the total sample (*N* = 716) in order to represent the proportion of the total sample affected.

Based on two engagements (B4 + A4), 29.9% of 716 crews (N = 214) were flagged as candidates for early elimination. Of these 214 crews, 200 actually failed to Q1, for a predictive accuracy of 93.5%. The 214 crews flagged for early elimination on the basis of two engagements produced a 10-engagement-based TTVIII mean score of 409.8. Fourteen crews (2% of the total sample) were misidentified on the basis of two engagements. That is, these 14 crews got off to a bad start on the two target engagements, yet managed to turn in superior performances on other engagements and eventually fired Q1 in spite of the contrary prediction.

With a three-predictor subset (B4 + A4 + A1), over a third of all crews (262 out of 716, or 36.6%) were identified for early elimination. Of the 262 identified crews, 247 (94.3%) actually failed to Q1. With four predictors (B4 + A4 + A1 + A2), almost half of all crews (47.3%) were flagged for early elimination, and the accuracy of the prediction was 92.6%. Prediction accuracy for subset sizes two through seven averaged 94.7%. Accuracy was 100% for subset sizes 8 and 9, but these figures were slightly inflated because crews were mathematically eliminated from Q1 with less than 500 and 600 accumulated points, based on eight and nine completed engagements, respectively.

Early Identification of Q1 Crews

The converse of early elimination is the early identification of crews with a high probability of firing Q1 on TTVIII. These crews could be flagged for early qualification awards and allowed to skip subsequent engagements, thereby saving ammunition in the process.

Formalizing early qualification predictions. For any given subset size, the minimum score (E_{qual}) necessary for early qualification can be predicted from an adaptation of the general equation defined earlier:

Equation 6: $$E_{\text{Qual}} = ([700 + (1.65 * SE)] / 10)* N_{\text{sub}}$$

Crews scoring at or above the specified scores could be pulled from the firing lane and awarded early Q1 status with 95% confidence that had they been allowed to fire all 10 TTVIII engagements, they would have received a score of 700 or greater. Table 11 presents the required E_{qual} score for each subset size. After completing the number of engagements listed in the Subset Size column, crews achieving a cumulative score equal to or greater than the corresponding value in the Minimum E_{qual} column would be eligible for early Q1 status.

Testing the early qualification model. Early qualification predictions for subset sizes from two through nine were tested with the 716 cases available in the database. Engagements in all subset sizes were based on optimal subsets of predictors as identified by stepwise regression procedures (see Table 3). For example, using a subset of two, each crew's scores on engagements B4 and A4 were summed. Sums were then partitioned into those < 185 and those ≥ 185. Crews with scores ≥ 185 were identified as early first-run qualifiers. This procedure was then repeated for each subset size. With

three engagements, for example, crews with scores of 268 or higher were flagged as early qualifiers. With four engagements, a summed score of 346 was required, and so on.

Table 11
Minimum E_{qual} Values for Early Q1 Identification

Subset Size	Prediction Equation	Minimum E_{qual}
2	$E_{qual} = [700 + (1.65 * 136.2)] / 10 * 2$	185
3	$E_{qual} = [700 + (1.65 * 116.5)] / 10 * 3$	268
4	$E_{qual} = [700 + (1.65 * 99.5)] / 10 * 4$	346
5	$E_{qual} = [700 + (1.65 * 88.5)] / 10 * 5$	423
6	$E_{qual} = [700 + (1.65 * 78.4)] / 10 * 6$	497
7	$E_{qual} = [700 + (1.65 * 67.2)] / 10 * 7$	567
8	$E_{qual} = [700 + (1.65 * 54.4)] / 10 * 8$	632
9	$E_{qual} = [700 + (1.65 * 39.0)] / 10 * 9$	688

For each subset size, the accuracy of these predictions was then tested by noting whether crews scoring at or above the stipulated cutoff points actually achieved TTVIII Q1 status, based on all 10 engagements. The pertinent question was what proportion of the crews identified by this procedure as eligible for early qualification awards actually qualified on their first-run. Results of this test are shown in Table 12.

Table 12
Identification of Early Qualifiers

Subset Size	Actual TTVIII Score ≥ 700	Actual TTVIII Score < 700	Predictive Accuracy
2	N = 143 (20.0%) Mean = 821.4	N = 40 (5.6%) Mean = 613.9	143/183 = 78.1%
3	N = 109 (15.2%) Mean = 840.0	N = 13 (1.8%) Mean = 639.7	109/122 = 89.3%
4	N = 82 (11.5%) Mean = 862.9	N = 4 (0.6%) Mean = 680.5	82/86 = 95.3%
5	N = 62 (8.7%) Mean = 881.9	N = 2 (0.3%) Mean = 659.5	62/64 = 96.9%
6	N = 77 (10.8%) Mean = 877.3	N = 3 (0.4%) Mean = 655.7	77/80 = 96.3%
7	N = 102 (14.2%) Mean = 864.9	N = 2 (0.3%) Mean = 661.0	102/104 = 98.1%
8	N = 138 (19.3%) Mean = 850.9	N = 5 (0.7%) Mean = 669.4	138/143 = 96.5%
9	N = 185 (25.8%) Mean = 833.1	N = 2 (0.3%) Mean = 693.0	185/187 = 98.9%

Predictive accuracy with this model was expected to cluster around 95%, and this was the case except for the two smallest subset sizes, where predictive accuracy was less. Markedly skewed predictor distributions would produce such diminished predictive accuracy. An examination of the data confirmed this to be the case. Non-normality was caused by a ceiling effect with N = 2 and N = 3 subset sizes. With N = 2, for example, early identification was predicated on a B4 + A4 score ≥ 185. Of the 183 cases with scores ≥ 185, 114 of them (62.3%) had a score of 200, the maximum possible. When the subset size was increased to three, 29.5% of crews had a score of 300, the maximum possible. In contrast, maximum possible scores were obtained by an average of only 3.6% of crews in subset sizes four through nine.

Because of the relatively low predictive accuracy with subset sizes two and three, the minimum recommended subset size for early identification of Q1 crews is $N = 4$. Based on four engagements, 86 out of 716 crews (12.0%) were flagged as early qualifiers. Of these 86 crews, 82 achieved a TTVIII total score of ≥ 700, thereby supporting the accuracy of the prediction. The mean score of this group was 862.9. Only 4 out of 86 identified crews failed to Q1, and even though these 4 crews fell short of the required 700 points for Q1 status, their mean score was 680.5. Predictive accuracy of the early qualification model exceeded 95% at every subset size from N = 4 through 9.

Combining Early Elimination with Early Identification

The combination of early elimination and early identification of Q1 crews is illustrated in the hypothetical outcome matrix of Table 13. This table is designed to illustrate the proportion of crews that could be recalled to the starting line and removed from the range after firing the number of engagements specified in the first column. No crews would be recalled after firing one engagement (the first row in the table). For subset sizes two and three, crews would be recalled only for early elimination (because of the relatively low predictive accuracy of early Q1 predictions for these two subset sizes).

Table 13
Combined Effect of Early Elimination and Early Identification of Q1 Crews

Subset Size	Minimum Score to Avoid Early Elimination	Minimum Score for Early Qualification	Predicted Early Elimination (%)	Predicted Early Qualification (%)	Total Crews Eliminated (%)	Prediction Accuracy (%)
1 (B4)	na	na	na	na	na	na
2 (A4)	95	na	29.9	na	29.9	93.5
3 (A1)	152	na	36.6	na	36.6	94.3
4 (A2)	214	346	47.3	12.0	59.4	93.2
5 (B3)	277	423	48.3	8.9	57.3	93.9
6 (B2)	342	497	46.9	11.2	58.1	96.2
7 (A5)	412	567	47.8	14.5	62.3	98.2
8 (B1)	500[a]	632	46.4	20.0	66.3	98.9
9 (B5)	600[a]	688	48.3	26.1	74.4	99.6

[a] Mathematical elimination.

From the second row of Table 13 it can be seen that after firing B4 and A4, the 29.9% of crews failing to accumulate at least 95 points would be recalled to the starting line and sent for remedial device-based training. After firing three engagements, a minimum of 152 points would be required to avoid early elimination. About 36% of the crews failed to meet this cutoff. Again, no early identification of Q1 crews would be made, because of the relatively low predictive accuracy associated with only three engagements.

Beginning with predictive subsets of size $N = 4$, crews could be recalled to the starting line for either early elimination or early qualification. After firing four engagements, for instance, 47.3% of crews in the database could have been recalled to the starting line because of failure to accumulate at least 214 points. Another 12% of crews could have been recalled and awarded early first-run qualification based on a score of at least 346 points. The combination of early elimination and early qualification would result in the removal of 59.4% of all crews from the firing lane based on 4 engagements.

Based on these results, we conclude that it is indeed possible to reduce the number of live-fire tank gunnery engagements without compromising the validity of the TTVIII evaluation process. Through use of the above-described target engagement reduction methodology, to include the specific guidance provided on how to select predictive engagement subsets, the ARNG can now conduct more resource-efficient live-fire tank gunnery evaluation without compromising the integrity of the process. Later on in the report, we will identify the approximate extent and kind of resource savings that can be expected.

Experiment 2

Encouraged by the above findings, we proceeded to test out our ARNG TTVIII target engagement reduction methodology on the AC. In general, our objective was to determine if this methodology would generalize to the AC without sacrificing the validity of the tank gunnery evaluation process.

Method

Data Source

Our data set consisted of first-run tank gunnery scores from 834 AC armor crews that fired TTVIII at Grafenwoehr, Germany, during 1993 and 1994.

Procedure

Stepwise multiple regression algorithms (SPSS, 1993, 1994) were used to determine if subsets of TTVIII engagements could be used to predict AC tank crews' TTVIII total scores. The best subsets of from two to nine engagements were identified and the

predictive validity specified for each after cross-validation was performed to determine the internal consistency and generalizability of the data.

Results

Descriptive Data

Table 14 compares ARNG and AC TTVIII data. AC mean scores were higher than ARNG mean scores, $t = 35.62, p < .0001$, and variances were lower. The lower variances can be understood by examining Tables 15 and 16. Negative skews are evident in both data sets, but the pattern is more pronounced among AC crews where almost all crews (97.7%) scored at least 700 and, therefore, qualified, on their first run. Moreover, perfect scores were attained by more than half of AC crews on all but two engagements. Relative to ARNG scores, AC scores are clustered toward the high end of the TTVIII scale, thereby restricting both variance and range. The lowest AC score was 475, vs. an ARNG low of 37.

Table 14
ARNG vs. AC TTVIII Data

	ARNG Data (N=716)		AC Data (N=834)	
	Mean	*SD*	Mean	*SD*
Total	614.0	194.0	891.5	82.1
A1	48.3	41.5	84.3	27.2
A2	45.4	41.1	77.2	32.4
A3	57.2	36.3	89.1	22.2
A4	58.3	41.6	88.7	24.0
A5	65.3	40.2	93.1	19.1
B1	78.4	38.7	96.1	15.4
B2	62.8	41.4	88.1	24.1
B3	55.6	35.4	90.6	17.5
B4	65.8	40.7	89.5	23.8
B5	76.4	40.2	94.4	19.1

Table 15
Measures of TTVIII Central Tendency

Measure	ARNG Data	AC Data
Mean	614.0	891.5
Median	642	906
Mode	759	1,000

In spite of the different levels of performance found between ARNG and AC crews, scores from both groups revealed similar patterns of relative performance on individual engagements. That is, engagements that were difficult (or easy) for AC crews were also difficult (or easy) for their ARNG counterparts. The corresponding patterns of relative performance are evident when mean engagement scores, and the percentages of perfect engagement scores, are rank ordered separately for ARNG and AC crews and then

compared (see Table 17). The rank ordering of mean engagement scores, as well as that of the percentage of crews firing a perfect score on individual engagements, were both similar, with r (Spearman) = .81, $p < .005$, and .90, $p < .001$, respectively.

Table 16

TTVIII Statistics for ARNG and AC Crews

Variable	ARNG Data	AC Data
Range (Total Score)	37 - 997	475 - 1,000
% of Crews ≥ 700 (Q1)	41.1	97.7
% Perfect Scores: Total	0.0	3.6
% Perfect Scores: A1	20.7	59.7
% Perfect Scores: A2	19.7	45.4
% Perfect Scores: A3	15.5	59.4
% Perfect Scores: A4	30.6	67.6
% Perfect Scores: A5	39.8	76.3
% Perfect Scores: B1	69.1	93.3
% Perfect Scores: B2	26.1	58.9
% Perfect Scores: B3	7.0	49.5
% Perfect Scores: B4	41.1	72.8
% Perfect Scores: B5	66.8	86.7

Table 17

Rank-Order Correspondence of TTVIII Engagement Performance for AC and ARNG Tank Crews

	Mean Engagement Score Rank		% Perfect Scores Ranked High to Low	
Engagement	ARNG Crews	AC Crews	ARNG Crews	AC Crews
A1	9	9	7	6
A2	10	10	8	10
A3	7	6	9	7
A4	6	7	5	5
A5	4	3	4	3
B1	1	1	1	1
B2	5	8	6	8
B3	8	4	10	9
B4	3	5	3	4
B5	2	2	2	2

The first row in the AC data correlation matrix (see Table 18) gives part-whole coefficients of correlation between the TTVIII total score and each individual engagement score. Other rows in the matrix present engagement score intercorrelations.

Part-whole correlations ranged from .241 (B1) to .494 (A2) with a mean of .352. Intercorrelations among engagements ranged from -.036 (B1, B4) to .095 (A4, B3) with a mean of .031. Part-whole correlation and predictor intercorrelation highlights for ARNG and AC crews are summarized in Table 19.

Table 18

TTVIII Correlation Matrix for AC Data

	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5
Total	.405	.494	.382	.433	.312	.241	.303	.328	.368	.251
A1		.055	.053	.058	.029	.001	-.001	.017	.030	.008
A2			.048	.091	-.004	.093	.018	.067	.030	-.011
A3				.074	.059	.021	-.008	.067	.076	.007
A4					.062	.007	.043	.095	.075	-.016
A5						.026	-.007	052	.055	.032
B1							-.016	.006	-.036	.071
B2								.040	-.013	-.034
B3									.028	.016
B4										.027

Table 19

Part-Whole Correlation and Predictor Intercorrelation Highlights for ARNG and AC Crews

	ARNG Crews	AC Crews
Part-whole correlations		
Range	.432 to .568	.241 to .494
Mean part-whole	.487	.352
Best predictor	B4 (.568)	A2 (.494)
Weakest predictor	B1 (.432)	B1 (.241)
Predictor intercorrelations		
Range	.076 to .323	-.036 to .095
Mean intercorrelation	.155	.031

The ARNG and AC data sets were similar in that relatively robust part-whole correlations were paired with relatively low intercorrelations among engagements. The low intercorrelation among engagements indicates that performance on one engagement cannot be predicted on the basis of performance on any other engagement. The relatively robust part-whole correlations, in contrast, indicate that every engagement has the potential of making its own contribution to total score predictions. The data sets differed, however, in that mean part-whole correlations and mean individual engagement intercorrelations were significantly attenuated among AC crews, relative to ARNG crews, $z = 3.20, p < .01$ and $z = 2.45, p < .05$, respectively. This attenuation may have been due to reduced score ranges and restricted variance in the AC data.

Split-Half Cross-Validation

A split-group, cross-validation design, similar to that applied to the ARNG data, was used to test for internal consistency and generalizability of the AC data. Half of the 834 tank crews were assigned at random (by SPSS Version 6.1 for Windows) to the normative groups and the other half were assigned to the cross-validation group. A series of least squares multiple regression prediction equations was developed based on the AC normative group. Stepwise procedures were used to select optimal subsets of 1, 2, 3, 4, 5,

6, 7, 8, and 9 predictor variables with a separate equation developed for each subset size. All prediction equations were statistically significant, producing Multiple *R's* ranging from .48 (based on one predictor) to .98 (based on nine predictors) and F ratios ranging from 125.13 ($df = 1, 415$) to 1,229.63 ($df = 9, 407$), with a rejection region of .0001 used for all equations.

The equations for the normative group were then tested on the cross-validation group and the accuracy of predictions for the two groups was compared. Results revealed that, regardless of the number of predictors involved, models developed from normative group data accounted for a comparable amount of TTVIII total score variance in the cross-validation group for subset sizes of N = 1 through 8. With nine predictors, the normative group equation was statistically less accurate when tested on the cross-validation group. This Multiple *R* difference (.982 vs. .972), however, was small enough to be of no practical value. Thus, the predictive models were found to be valid and, therefore, likely to maintain similar efficiency when used to predict the TTVIII total scores of other AC tank crew samples (at least those consisting of crews firing TTVIII in Grafenwoehr, Germany). Given the similar outcomes of the separate group analyses, along with our desire to obtain the best possible predictions form the largest sample size possible, subsequent analyses were conducted on pooled-group data (N = 834).

Development of Pooled-Group AC Prediction Equations

Using stepwise multiple regression routines described for the ARNG, prediction equations were developed for the best subsets of 1, 2, 3, 4, 5, 6, 7, 8, and 9 engagements fired by AC crews. Prediction equations for every subset size were statistically significant, producing Multiple *R's* ranging from .49 (based on one predictor) to .98 (based on nine predictors) and F ratios ranging from 268.64 ($df = 1, 832$) to 2,555.44 ($df = 9, 824$). Results for both ARNG and AC crews are summarized in Table 20, while the derived AC prediction equations are shown in Table 21.

The first four columns in Table 20 pertain to ARNG data, and the second four columns pertain to AC Data. Order of entry differed somewhat for ARNG and AC crews, but the sequences were similar in that the first four predictors, which accounted for a majority of TTVIII total score variance, were the same in both groups. The last two columns of Table 20 test for differences in predictive accuracy of ARNG vs. AC equations at each subset size. A significant outcome indicates that the Multiple *R's* for the two groups differed reliably. Predictive accuracy was lower among AC crews for subset sizes one through six, whereas no differences were found for subset sizes seven and eight. For nine predictors, the AC model was more effective than the ARNG model, although the difference was small enough to be of no practical value.

Table 20

Stepwise Multiple Regression Results for ARNG and AC Crews

ARNG Crews				AC Crews				ARNG vs. AC	
Order of Entry	Multiple R	Adj. R^2	Standard Error	Order of Entry	Multiple R	Adj. R^2	Standard Error	z	p
1 B4	.568	.322	159.80	1 A2	.494	.243	71.46	1.97	< .05
2 A4	.713	.507	136.23	2 A4	.630	.395	63.89	3.00	< .01
3 A1	.801	.640	116.47	3 A1	.724	.523	56.75	3.59	< .01
4 A2	.859	.737	99.53	4 B4	.791	.623	50.41	4.22	< .01
5 B3	.891	.792	88.51	5 A3	.844	.710	44.22	3.80	< .01
6 B2	.915	.832	78.42	6 B2	.892	.793	37.34	2.53	< .05
7 A5	.939	.880	67.21	7 B5	.928	.860	30.76	1.51	ns
8 B1	.960	.921	54.37	8 A5	.960	.920	23.21	< 1	ns
9 B5	.981	.962	39.03	9 B3	.983	.965	15.36	-2.00	< .05

Table 21

Prediction Equations for Subset Sizes 1 to 9 (AC Crews)

Subset Size	Prediction Equation
1	Y′ = 794.7724 + 1.2531(A2)
2	Y′ = 682.7070 + 1.1628(A2) + 1.3422(A4)
3	Y′ = 600.6519 + 1.0843(A1) + 1.1169(A2) + 1.2764(A4)
4	Y′ = 512.4405 + 1.0607(A1) + 1.0994(A2) + 1.1989(A4) + 1.1003(B4)
5	Y′ = 432.3853 + 1.0206(A1) + 1.0710(A2) + 1.0961(A3) + 1.1356(A4) + 1.0291(B4)
6	Y′ = 347.9194 + 1.0231(A1) + 1.0599(A2) + 1.1071(A3) + 1.0920(A4) + 0.9825(B2) + 1.0452(B4)
7	Y′ = 242.5341 + 1.0165(A1) + 1.0669(A2) + 1.1012(A3) + 1.1062(A4) + 1.0110(B2) + 1.0206(B4) + 1.1050(B5)
8	Y′ = 158.8839 + 0.9997(A1) + 1.0753(A2) + 1.0556(A3) + 1.0593(A4) + 1.0602(A5) + 1.0166 (B2) + 0.98174(B4) + 1.0728(B5)
9	Y′ = 88.4395 + 0.9971(A1) + 1.0450(A2) + 1.0123(A3) + 1.0008(A4) + 1.0207(A5) + 0.9898 (B2) + 0.9983(B3) + 0.9717(B4) + 1.0574((B5)

Predictive Accuracy and Number of Engagements

Table 22 summarizes the number of TTVIII engagements needed for various levels of predictive accuracy for both ARNG and AC crews. The table is based on the best possible combinations of engagements, as determined by stepwise multiple regression procedures. Seven engagements are sufficient to ensure predictive accuracy of ≥ 85% with either ARNG or AC crews.

Table 22
Relationship Between TTVIII Predictive Accuracy and Required Number of Engagements for ARNG and AC Crews

Predictive Accuracy	No. of Engagements ARNG Crews	No. of Engagements AC Crews
100%	10	10
95%	9	9
90%	8	8
85%	7	7
80%[a]	6	6
70%	4	5

[a] actually .793.

Random Subsets of Engagements

The predictive accuracy of randomly selected subsets of engagements was tested on the AC data set at each subset size from two through nine. Five randomly selected combinations of engagements were tested at each subset size. In all, 40 random subsets were used, 5 at each of 8 possible subset sizes. The manner in which the random subsets were constructed is described in the *Random Subsets of Engagements* section of Experiment 1. The same random subsets were used in both experiments.

For each of the 40 random subsets, multiple regression procedures were used to construct prediction equations. For each subset size, the predictive power of random subsets of engagements was compared to the predictive power of the best possible combination of engagements as determined by multiple regression procedures.

At every subset size (from two through nine) z tests between the mean Multiple R for the random subsets and the Multiple R for the best predictors indicated that the best predictors were superior. Z scores were 4.92, 5.10, 4.39, 5.70, 6.56, 5.99, 7.54, and 5.93 for subset sizes two through nine, respectively. All z values were significant at $p < .01$. Details of the 40 multiple regression analyses are given in Appendix C, and a summary of the results is presented in Table 23 in order to contrast the relative magnitudes of Multiple R and R^2 values for random subsets vs. the best subsets. From Table 23, it can be seen that substantial differences in predictive power occurred at subset sizes two through six. With larger subsets sizes (seven through nine), differences in predictive power were less pronounced, but the differences were statistically significant nonetheless.

Table 23 results contrast with those obtained from the ARNG data sample. With ARNG crews, the best subsets of predictors (as determined by regression procedures) were superior to random subsets of predictors only up to subset size six. Randomly constituted subsets of seven, eight, or nine predictors were as effective as the best subsets of corresponding size. For AC crews, however, the best predictors were superior at every subset size. This difference can probably be attributed to the extreme skew in the AC data. It will be recalled that the AC data set contained less variance than the ARNG set, due to the fact that 97.7% of crews scored ≥ 700 on a scale from 1 to 1,000. The AC data set also had lower part-whole correlations, possibly due to truncated ranges among both

predictor variables and the criterion. Thus, relative to the ARNG data, the AC data set had fewer "good" predictors (and hence more "poor" predictors). (See Table 19.) With fewer good predictors to go around, randomly constituted subsets of AC engagements were more susceptible to excluding one of the better predictors and more vulnerable to including one (or more) of the relatively poor predictors, thereby impairing the efficiency of random subsets and ensuring the superiority of the best subsets.

Table 23

Multiple R and R^2 Values for the Best Subset and for Random Subsets of Engagements (AC Data)

	Multiple *R*		Adjusted R^2	
Subset Size	Best Subset	Random Subsets	Best Subset	Random Subsets
2	.630	.462	.395	.213
3	.724	.583	.523	.340
4	.791	.695	.623	.485
5	.844	.743	.710	.552
6	.892	.804	.793	.645
7	.928	.875	.860	.764
8	.960	.919	.920	.845
9	.983	.969	.965	.938

An AC Shortcut Prediction Model

The shortcut prediction model that was developed and tested on the ARNG data set was also tested on the AC sample. It will be recalled that the shortcut model consists of three basic steps:

1: Add the engagement scores of the desired subset size
2: Divide the sum by N_{sub}, the number of engagements in the subset
3: Multiply the quotient by 10

In this manner, each engagement is weighted equally (by dividing by N) and the mean of all engagements in the subset is extrapolated to a 10-engagement TTVIII total score (by multiplying by 10).

The efficacy of the shortcut prediction model for AC crews was tested by constructing a series of shortcut predictor variables. For each subset size (from $N = 2$ to 9), six shortcut predictor variables were constructed. The first shortcut variable for each subset size was based on the best set of engagements identified in the AC stepwise regression procedures (see the right-hand side of Table 20). For example, for the $N = 2$ subset, the first shortcut predictor variable was calculated by the following procedure:

$$[(A2 + A4)/2] \times 10$$

The resulting shortcut score was then used as an independent variable to predict actual TTVIII scores.

The other five shortcut predictor variables (at each subset size) were based on the randomly constituted engagement subsets described earlier. These subsets are listed in Appendix C. The first random shortcut predictor variable, for example, was created with the following procedure:

$$[(A3 + B5)/2] \times 10$$

Each random shortcut predictor variable was then used to predict actual TTVIII scores.

The results of the shortcut test appear in Table 24. The first column under the "Full Regression Models" heading shows R^2 values at each subset size for the best engagement predictors as determined by stepwise multiple regression procedures. The second column under the Full Regression Models heading shows mean R^2 values from five randomly constituted subsets of engagements. The data in the two columns under Full Regression Models were adapted from Table 23 and are reproduced here to facilitate comparisons with the shortcut-based prediction model outcomes.

Table 24
R^2 Values for Full Regression Models Vs Shortcut Regression Models (AC Data)

	Full Regression Models (R^2)		Shortcut Regression Models (R^2)	
Subset Size	Best Subset	Random Subsets	Best Subset	Random Subsets
2	.395	.213	.394	.211
3	.523	.340	.522	.337
4	.623	.485	.624	.485
5	.710	.552	.711	.550
6	.793	.645	.794	.644
7	.860	.764	.860	.764
8	.920	.845	.920	.843
9	.965	.938	.965	.938

The last two data columns in Table 24 are based on shortcut regression models. The next-to-last column presents R^2 values using shortcut models based on the best predictors for each subset size. The last column contains mean R^2 values obtained from five shortcut regression models based upon randomly selected subsets of predictors. The results indicate that the shortcut method can be used successfully with reduced subsets of any size. Table 23 also reinforces the earlier finding that for the AC data set random subsets of predictors do not work as well as the best subsets, and this is the case for both full regression and shortcut regression methods.

Early Elimination and Early Qualification of AC Crews

Because the vast majority of AC crews achieved Q1, development of early elimination and early qualification procedures, like those developed for the ARNG data, were unnecessary. With the high Q1 rate found, a prediction of early qualification could be applied to every crew in the AC data set with an accuracy rate of 97.7%.

Discussion

ARNG and AC Similarities and Differences

Similarities. Both the ARNG and AC data sets were internally consistent, as revealed by split-half cross-validation procedures. Hence the results from both data sets have potential generalizability. The ARNG results are probably more generalizable because of the variety of units contained in the SIMITAR database. For the AC, generalizability will depend on how representative Grafenwoehr-firing units are of armor units stationed stateside.

Both data sets revealed relatively low intercorrelations among engagement scores and relatively robust part-whole correlations. And in spite of the different levels of performance found between ARNG and AC crews, scores from both groups revealed similar patterns of relative performance on individual engagements. That is, engagements that were difficult (or easy) for AC crews were also difficult (or easy) for their ARNG counterparts.

Perhaps of greatest importance, subsets of engagements proved to be effective predictors of TTVIII total scores for both the AC and ARNG, and shortcut prediction methods worked well for both AC and ARNG crews. Thus, subsets of engagements can be used to predict TTVIII total scores among both ARNG and AC crews with known degrees of predictive accuracy.

Differences. Despite the fundamental similarities existing between the ARNG and AC data, differences were found. The most striking of which was found between mean TTVIII scores. On the average, AC crews scored 277.5 points higher than ARNG crews. These consistently high scores resulted in 97.7% of AC crews attaining Q1, vs. 41.1% of ARNG crews. While these performance differences were not surprising given the vastly greater training time available to AC units (Eisley & Viner, 1989), the elevated AC test scores also had the effect of producing reduced variance and restricted score ranges. It was to be expected that reduced variance would suppress part-whole correlations and impair the effectiveness of regression-based prediction equations, but it also produced the more subtle effect of impairing the effectiveness of randomly selected subsets of engagements. For AC crews, at every subset size, randomly selected subsets of engagements failed to work as well as the best subsets determined by multiple regression procedures. In contrast, for ARNG crews randomly selected subsets of engagements worked just as well as optimized subsets, as long as at least 7 engagements were used. Elevated AC test scores also precluded the necessity of developing early elimination and early qualification predictions. Although these procedures promise substantial resource efficiencies among ARNG crews, they were not applicable to AC crews.

Because of the subtle but important differences between the ARNG and AC data sets, training implications are somewhat different. For this reason, our discussion will focus first on ARNG units and, then, on a separate consideration of AC training implications.

Resource-Efficient Tank Gunnery Evaluation in the ARNG

The findings of this research reveal that more resource-efficient evaluation of tank gunnery proficiency in ARNG armor units is possible by reducing the number of engagements fired on TTVIII. Fewer engagements can be fired, and then the scores on these engagements can be used to predict a 10-engagement-based TTVIII total score. While elimination of even one engagement results in some loss of predictive precision (albeit small), the extent of this loss can now be specified. In fact, it is now possible to specify how much loss in predictive precision is associated with dropping any given number of engagements from TTVIII (see Table 22). Thus, a user of this target engagement reduction methodology can now stipulate the level of predictive accuracy desired and then determine the engagement subset size associated with that level of precision.

Specific vs. random subsets of engagements. Our ARNG findings also suggest which TTVIII engagements should be fired for each subset size (from one to nine). For subsets ranging in size from one to six engagements, it is important to use the specific engagements identified by multiple regression statistical routines. For subsets containing seven to nine engagements, however, specific engagements matter very little. Seven engagements selected at random, for example, will work as well as the best seven statistically-identified predictive engagements.

Practical implications. If only nine engagements are to be fired for the sake of resource efficiency, then any engagement can be randomly eliminated. The same is true when the number of engagements is reduced to eight, or even seven. Thus, up to three engagements can be randomly selected and dropped with little concern for which specific engagements they are. The random selection process can take place after the conclusion of tank gunnery training. In this way, not only is TTVIII shortened, but units are precluded from concentrating their training on only those engagements that are to be evaluated later on TTVIII. Thus, training could proceed as if all 10 engagements were going to be fired. Then, as many as three engagements could be selected at the last minute for exclusion from the table.

The Shortcut Prediction Model for ARNG Tank Crews

From the standpoint of implementation, one of the more important products of this research is the shortcut prediction model. By using this model, it is possible to fire a reduced-engagement version of TTVIII, use the results to estimate 10-engagement-based TTVIII scores, and never use any computational procedures more complicated than simple arithmetic. The shortcut prediction model consists of selecting a subset of engagements upon which a TTVIII total score prediction is to be based, firing the selected subset of engagements, adding the individual engagement scores, dividing the sum by the number of engagements in the predictive subset, and then multiplying by 10. The result is a predicted 10-engagement-based TTVIII total score, the accuracy of which

will differ little from the accuracy of a prediction based on a more complex multiple-regression-based prediction equation.

Early Elimination and Qualification Predictions for ARNG Crews

All 10 TTVIII engagements are useful predictors, in the sense that every engagement accounts for a statistically significant degree of unique variance in TTVIII total scores. Some engagements, however, account for more variance than others and, hence, are better predictors. By administering the most predictive engagements early in the evaluative process, it is possible to use a small subset of predictors to identify crews with little chance of firing Q1. Conversely, it is also possible to identify crews with a high probability of firing Q1, based on the same subset of key engagements. For example, after firing four engagements (B4, A4, A1, and A2), crews with less than 214 cumulative points have no more than a 5% probability of firing Q1. Moreover, crews receiving at least 346 cumulative points on the same 4 engagements have at least a 95% probability of firing Q1. Early elimination and early qualification predictions should be based only on statistically identified engagements. If such predictions are to be based upon four engagements, for instance, then they should be B4, A4, A1, and A2. They can be fired in any order.

Early elimination predictions. Accurate early elimination predictions can be based on as few as two specific engagements (B4 and A4). Based only on these two engagements, it was possible to identify 29.9% of 716 crews ($N = 214$) that had little chance of firing Q1. The accuracy of this prediction was 93.5%. That is, 200 of the 214 identified crews actually failed to fire Q1. When a third engagement was added to the predictive subset (B4, A4, and A1), the proportion of identified crews rose to 36.6%, with a predictive accuracy of 94.3%. With four predictors (B4, A4, A1, and A2), 339 out of 716 crews (47.3%) were predicted not to Q1, and 314 of the 339 (92.6%) actually failed to Q1.

Early qualification predictions. Accurate early qualification predictions require the use of all four of the above engagements (B4, A4, A1, and A2). Based on these four engagements, it was possible to identify 86 out of 716 crews (12.1%) with a high probability of achieving Q1 status. Of the 86 identified crews, 82 (95.3%) actually fired Q1.

Early elimination and early qualification predictions can be used in tandem. Based on four engagements, for instance, 59.4% of all crews in the ARNG database were flagged for either early elimination or early qualification. Thus, for 6 out of 10 crews, Q1 status was predictable (with approximately 95% predictive accuracy) after they fired only four engagements. After this point, Q1 outcome on TTVIII was in question for only the remaining 4 out of 10 crews. Early elimination and early qualification predictions have implications for potential resource efficiencies, as discussed below in the Projected TTVIII Resource Efficiencies section.

Resource-Efficient Tank Gunnery Evaluation in the AC

Our findings also reveal that more resource-efficient evaluation of tank gunnery proficiency in AC armor units is possible by reducing the number of engagements fired on TTVIII. Fewer engagements can be fired and, then, the scores on these engagements can be used to predict a 10-engagement-based TTVIII total score. The predictive accuracy, however, of subsets ranging in size from one to six is less among AC crews than it is among ARNG crews. When the predictive subset size reaches seven or eight engagements, predictive accuracy is equal for AC and ARNG crews. Moreover, nine-engagement-based predictions are slightly more precise based on the AC data sample (see Table 20). Generally, it would seem unwise to base AC TTVIII total score predictions on less than seven engagements.

Specific subsets vs. randomly selected subsets of engagements. At all subset sizes, greater predictive accuracy for AC crews is achieved by using subsets consisting of the best predictors. Using specifically-identified predictors is especially important if predictions are to be based on six or fewer engagements. Table 20 provides guidance on which engagements to include at each subset size and Table 23 indicates the sacrifice in predictive accuracy that is to be expected by using randomly constituted subsets vs. the best subsets of engagements.

Practical implications. If, for the sake of resource efficiency, only nine engagements were to be fired, the best tactic would be to drop engagement (B1). This engagement has been identified by regression procedures as contributing the least amount of incremental unique variance. Dropping any other engagement will result in a (statistically significant) loss of predictive accuracy. Reference to Table 23 indicates that elimination of the regression-determined engagement B1 will result in a 10-engagement-based TTVIII total score prediction that incorporates 96.5% predictive accuracy. Elimination of a randomly selected engagement, on the other hand, will result in a total score prediction that incorporates 93.8% predictive accuracy. The difference between 96.5% and 93.8% is only 2.7%, but it is statistically significant. As to whether this difference is practically significant depends upon the judgment of the individual user.

The sacrifice in predictive accuracy resulting from random elimination grows, however, as the size of the predictive subset shrinks. If the number of engagements is reduced to eight, the discrepancy in predictive accuracy between the best predictors and randomly selected subsets increases to 7.5%. With seven engagements, the discrepancy between the two selection procedures is 9.6%. Thus, resource efficiencies through a reduction in the number of TTVIII engagements in AC units should proceed only when close attention is paid to the selection of specific engagements.

The Shortcut Prediction Model for AC Tank Crews

The shortcut prediction model was as successful among AC crews as it was among ARNG crews. By using this simple computational model, it is possible to fire a reduced-engagement version of TTVIII, use the results to estimate 10-engagement-based TTVIII

scores without resorting to anything more complicated than simple arithmetic, and obtain as much predictive accuracy as if a more complex multiple-regression-based prediction equation had been used.

AC Early Elimination and Early Qualification Predictions

The significant difference between ARNG and AC mean first-run TTVIII total scores was accompanied by a marked difference in the proportion of ARNG and AC crews achieving Q1. (Among ARNG crews, the figure was 41.1%. Among AC crews, the figure was 97.7%.) Accordingly, with such a high Q1 level in the AC sample, early elimination and early qualification predictions for AC crews are unnecessary. That is because qualification outcomes can be predicted with 97.7% accuracy before a single engagement is fired. With no other information about a particular crew other than that it is from the AC sample, one could guess that the TTVIII outcome will be Q1 and enjoy predictive success 97.7% of the time. Thus, the number of predicted early eliminations will be negligible and virtually every crew will be a candidate for early qualification. Of course, if almost 98% of crews Q1, one has to question the need to fire TTVIII at all, at least in its present form. Presumably, resources could be used in other ways, such as on TTXII, or perhaps on more difficult TTVIII-type engagements in order to expand crew-level gunnery capabilities. In fact, TTVIII engagements have recently been modified to include as many as four targets on some engagements (Department of the Army, 1998).

Projected TTVIII Resource Efficiencies

In the following sections, resource efficiencies attributed to an across-the-board reduction in the number of TTVIII engagements apply to either the ARNG or AC. Resource efficiencies attributed to implementation of early qualification or early elimination procedures, however, apply only to the ARNG.

Resource efficiencies can be realized in three ways from implementation of the findings of this research: through (1) an across-the-board reduction in the number of TTVIII engagements, (2) implementation of early qualification procedures, and (3) implementation of early elimination procedures. Resource efficiencies from an across-the-board reduction in engagements and from implementation of early qualification procedures are straightforward and relatively easy to estimate. Resource efficiencies from early elimination procedures, however, are more difficult to quantify.

Resource efficiencies from an across-the-board reduction in the number of TTVIII engagements fired. If fewer TTVIII engagements were fired, then fewer rounds of ammunition would be needed. A reduction in the number of engagements from 10 to 7, for example, would result in approximately a 30% across-the-board savings in ammunition. Of course, there should be other savings as well, including reduced tank operating (OPTEMPO) costs, but these savings are difficult to quantify because it is impossible from our perspective to anticipate how crews would spend the extra time saved by not firing 3 of the 10 TTVIII engagements.

Resource efficiencies from implementation of early qualification procedures. Potential savings from implementation of early qualification procedures are also relatively easy to estimate. After firing four engagements, for example, approximately 12% of crews can be identified as having at least a 95% probability of firing Q1. Ostensibly, these crews could be recalled to the starting line and removed from the range, thereby conserving ammunition that would have been fired on engagements 5 through 10 (or 5 through 7, if a reduced subset strategy were in place). This would result in 12% of crews firing 60% fewer rounds (based on skipping engagements 5 through 10), for a net ammunition savings of 7.2%. This potential efficiency would be reduced to 3.6% if a reduced subset methodology of seven engagements were in place (12% of crews times 30% of engagements = 3.6% ammunition savings). The 3.6% savings would be incremental to across-the-board efficiencies resulting from reducing the number of engagements to 7 from the current 10. Thus, ammunition savings from an across-the-board reduction in the number of TTVIII engagements (from 10 to 7) and from implementation of early qualification procedures would amount to approximately a 33.6% total ammunition savings.

Resource efficiencies from implementation of early elimination procedures. Potential savings from early elimination procedures are difficult to quantify, because of the different refire procedures used in ARNG units. Yet, potential economies of early elimination are hard to ignore because of the relatively large proportion of crews that can be identified on the basis of a small number of engagements. Perhaps the best way to think of resource efficiencies resulting from early elimination procedures is in the context of enhanced efficiency of range operation that would unquestionably redound to ARNG units. By identifying and removing relatively deficient crews, the range could be made more readily available to crews with a better chance of firing Q1. These more proficient crews should achieve qualification without the need for many reruns. Moreover, when the removed crews attained device-based training proficiency standards (see Hagman & Smith, 1996) and return to the range, they should then be able to rapidly achieve qualification.

Summary of resource efficiencies. It is estimated that 33.6% of current TTVIII ammunition costs could be saved by implementing an across-the-board reduction in the number of TTVIII engagements from 10 to 7, and by implementing an early qualification program wherein exceptionally proficient crews are pulled from the range and awarded special recognition after firing four engagements. These projected savings do not include projected enhanced range operating efficiencies from implementation of an early elimination program.

Summary and Recommendations

The findings of this research suggest that more resource-efficient live-fire tank gunnery evaluation is indeed possible for both the ARNG and AC without sacrificing the validity of the evaluation process. In support of this notion, we have (a) presented a target engagement reduction methodology developed to support resource-efficient, live-fire gunnery evaluation on TTVIII, (b) identified which specific target engagement

subsets to use for best results, and (c) estimated the magnitude of resource savings to be anticipated from use of these subsets for purposes of crew-level tank gunnery proficiency certification.

Although the specific structure and content of more resource-efficient live-fire evaluation scenarios will vary with the particular goals of the user (e.g., unit commanders), we suggest consideration of the following five-step scenario as one that will provide the best "readiness bang for the evaluation buck" for the ARNG.

Step 1. Fire a maximum of seven TTVIII engagements, with the first four being the statistically identified best predictors (B4 + A4 + A1 + A2) of TTVIII total scores.

Step 2. Use crew performance on these four best predictive engagements to support early qualification decisions, and crew performance on the first two of these engagements (B4 + A4) to support early elimination decisions. (Require device-based training for crews that are eliminated early [See Hagman & Morrison, 1996 for details])

Step 3. Add three engagements, selected *at random* from those remaining, to arrive at the desired total subset size of seven. Do this within a month of scheduled TTVIII firing to discourage "training to the test."

Step 4. Predict TTVIII scores from tank crew performance (i.e., calculated via the shortcut prediction model) on this seven-engagement subset.

Step 5. Use predicted scores from this subset, along with the early qualification scores from Step 2, to evaluate/certify crew-level tank gunnery proficiency on TTVIII.

We expect that ARNG adherence to these five steps will (a) produce an across-the-board reduction of three engagements, (b) enable implementation of early elimination/qualification procedures, and (c) support accurate TTVIII proficiency certification decisions--all at a substantial resource savings.

For the AC, the recommended scenario would involve only the following three steps:

Step 1. Fire the seven TTVIII engagement subset (i.e., A2, A4, A1, B4, A3, B2, and B5) found statistically to best predict total TTVIII scores.

Step 2. Predict TTVIII scores from tank crew performance (i.e., calculated via the shortcut prediction model) on this seven-engagement subset.

Step 3. Use predicted scores from this subset to evaluate/certify crew-level tank gunnery proficiency

We expect that AC adherence to these three steps will support accurate crew certification decisions at a substantial resource savings.

The confidence with which we offer these recommendations has been tempered somewhat because some TTVIII engagements have been changed (Department of the Army, 1998) since we began writing this report. Thus, additional research is needed to determine whether or not our findings still apply to this new set of engagements. Our target engagement reduction methodology, in contrast, should still apply regardless of the specific set of engagements upon which it is applied.

References

Department of the Army (1993). *Tank gunnery training (Abrams) (FM 17-12-1-2)*. Washington, DC: Author.

Department of the Army (1998). *Tank gunnery training (Abrams) (FM 17-12-1-2)*. Washington, DC: Author.

Eisley, M., & Viner, M.P. (1989). *Nationwide survey of soldier perceptions of Reserve Component training*. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A217 639)

Hagman, J. D., & Morrison, J. E. (1996, November-December). A strategy for efficient device-based tank gunnery training in the Army National Guard. *Armor*, 48-50

Hagman, J. D., & Smith, M. D. (1996). Device-Based prediction of tank gunnery performance. *Military Psychology, 8,* 59-68.

Krug, R. E., & Pickell, G. A. (1996, February). SIMITAR sharpens the guard. *Army*, 57-59.

McAndrews, K. (1997, April). The $743 Million Question. *National Guard,* 18-20, 24-26.

Shaler, M. D. (1995). *Compressed gunnery program for Abrams Tank battalions: Trainer's handbook*. Arlington, VA: Houston Associates.

Smith, M. D. (1998a). *Assessment of the SIMITAR gunnery training strategy through development of a database of gunnery outcome measures* (Research Rep. 1721). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A344 930)

Smith, M. D. (1998b). *User's manual for an Army National Guard armor and mechanized infantry gunnery training assessment database*. (Research Product 98-34). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A347 235)

SPSS, Inc. (1993). *SPSS base system syntax reference guide.* (Release 6.0). Chicago, IL: SPSS, Inc.

SPSS, Inc. (1994). *SPSS 6.1 for Windows update.* Chicago, IL: SPSS, Inc.

Tatsuoka, M.M. (1969). *Selected topics in advanced statistics: Validation studies.* Champaign, IL: The Institute for Personality and Ability Testing.

U.S. Army Armor School (1995). *Enhanced mounted brigade training strategy.* Fort Knox, KY: Author.

Appendix A
Characteristics of TTVIII Engagements

Engagement	Task	Conditions/Situation	Target(s)	Ammo
A1	Engage multiple targets (defense).	Move from turret-down to hull-down. Using GAS, BATTLESIGHT from a stationary tank. Computer and LRF failure	1 moving T-72, 900-1,300 m. 1 stationary T-72, 900-1,300m.	3 rds TPDS-T
A2	Engage simultaneous targets (defense).	Move from turret-down to hull-down. Using GPS, PRECISION.	1 stationary BMP, 900-1,100m.	2 rds HEAT-TP-T
		Using TC's sight from a stationary tank.	1 BTR, 800-1,000m.	50 rds Cal .50
A3	Engage multiple targets (offense).	Using GPS from a moving tank.	2 sets of troop targets, 400-600m and 700-900m.	200 rds 7.62mm
A4	Engage multiple targets (offense).	Using GPS, PRECISION from a moving tank. NBC environment.	2 stationary T-72s, 1,400-1,600m	3 rds TPDS-T
A5	Engage multiple targets (offense). (Swing task.)	Using GPS, PRECISION from a moving tank.	2 moving T-72s, 1,400-1,600m.	3 rds TPDS-T
A5A	Engage multiple targets (offense). (Alternate.)	Using GPS, PRECISION from a moving tank.	1 stationary T-72, 1 moving T-72. 1,400-1,600m	3 rds TPDS-T
B1	Engage a target (defense). (Swing task.)	Move from turret-down to hull-down. Using GPSE, PRECISION from a stationary tank. Three-man crew, loader is killed.	1 stationary T-72, 1,400-1,600m.	2 rds TPDS-T
B2	Engage multiple targets (defense).	Move from turret-down to hull-down. Using GPS, PRECISION from a stationary tank.	2 stationary BMPs, 1,200-1,400m.	3 rds HEAT-TP-T
B3	Engage multiple targets (offense).	Using GPS from a moving tank. NBC environment.	1 stationary BMP, 400-600m. 1 RPG team, 400-600m.	1 rd HEAT-TP-T 50 rds 7.62 mm
B4	Engage multiple targets (offense).	Using GPS, PRECISION from a moving tank.	1 stationary T-72, 1,300-1,500m. 1 moving T-72, 1,300-1,500m.	3 rds TPDS-T
B5	Engage a target (defense).	Move from turret-down to hull-down. Using GAS with illumination from a stationary tank. TIS failure.	1 stationary T-72, 1,200-1,400m.	2 rds TPDS-T
B5A	Engage a moving target (defense). (Alternate.)	Move from turret-down to hull-down. Using GPS, PRECISION from a stationary tank.	1 moving T-72, 1,700-1,900m.	2 rds TPDS-T

Note: Crews fire a total of 10 engagements. A5A and B5A are alternate engagements which can be fired in lieu of A5 and B5, respectively. Crews fire either the main engagement or its alternate, never both. When alternate engagements were fired, they were substituted for the main engagements.

Appendix B
Random Subsets of N = 2 Through 6 (ARNG Data)

Random subsets of $N = 2$. Table B-1 presents the results for subsets of $N = 2$. The first five rows present multiple regression results for the five random subsets. Means in the sixth line of the table are based upon the five individual random subsets. The cell under the "p" column for the "Mean" row is blank because it is meaningless to calculate a mean probability level. The last line in the table provides multiple regression results based upon the two best predictors (B4 + A4).

Table B-1
Random Subsets of N = 2 vs. the Two Best Predictors

Predictors	Multiple R	Adjusted R^2	$F(2, 713)$	p	SE
A3, B5	.615	.377	217.01	<.0001	153.16
A1, B2	.661	.435	275.96	<.0001	145.85
A4, B5	.658	.432	272.78	<.0001	146.22
B4, B5	.642	.411	250.62	<.0001	148.86
B1, B3	.600	.358	200.48	<.0001	155.42
Mean	.635	.403	243.37		149.90
Best 2	.713	.507	368.42	<.0001	136.23

Multiple *R's* for the random subsets ranged from .600 to .661, with an average of .635. Two-predictor random subsets accounted, on average, for 40.2% of criterion (TTVIII) variance and produced *SE*s of approximately 149.90 along with highly significant F values in excess of 200. By comparison, the two best predictors accounted for 50.7% of criterion variance. A test between the mean Multiple R for the random subsets and the Multiple R for the two best predictors indicated that the two best predictors were superior to random subsets, $z = 2.64, p < .01$.

Random subsets of $N = 3$. Table B-2 presents the results for subsets of $N = 3$. The last line in the table provides multiple regression results based upon the three best predictors (B4 + A4 + A1).

Table B-2
Random Subsets of N = 3 vs. the 3 Best Predictors

Predictors	Multiple R	Adjusted R^2	$F(3, 712)$	p	SE
A1, A4, B5	.764	.583	333.71	<.0001	125.33
B2, B3, B5	.709	.501	239.84	<.0001	137.10
A1, B3, B5	.716	.511	249.59	<.0001	135.72
B2, B3, B4	.722	.519	258.13	<.0001	134.55
A3,A4,B4	.775	.600	357.99	<.0001	122.75
Mean	.737	.542	287.85		131.09
Best 3	.801	.640	423.89	<.0001	116.47

Three-predictor random subsets accounted, on average, for 54.2% of criterion (TTVIII) variance and produced *SE*s of approximately 131.09 along with significant *F* values exceeding 200. By comparison, the three best predictors accounted for 64.0% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple *R* for the three best predictors indicated that the three best predictors were superior to random subsets, $z = 2.89, p < .01$.

Random subsets of N = 4. Table B-3 presents the results for subsets of $N = 4$. The last line in the table provides multiple regression results based upon the four best predictors (B4 + A4 + A1 + A2).

Table B-3
Random Subsets of N = 4 vs. the Four Best Predictors

Predictors	Multiple R	Adjusted R^2	$F(4, 711)$	p	SE
A2, A3, A5, B1	.769	.589	257.23	<.0001	124.36
A1, A2, A4, B3	.836	.697	411.29	<.0001	106.87
A5, B2, B4, B5	.808	.651	334.90	<.0001	114.55
A3, A5, B4, B5	.815	.662	350.37	<.0001	112.86
A1, A2, A4, B5	.838	.700	418.95	<.0001	106.18
Mean	.813	.660	354.55		112.96
Best 4	.859	.737	501.29	<.0001	99.53

Four-predictor random subsets accounted, on average, for 66.0% of criterion (TTVIII) variance and produced *SE*s of approximately 113 along with *F* ratios generally in excess of 200. By comparison, the 4 best predictors accounted for 73.7% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple *R* for the 4 best predictors indicated that the 4 best predictors were superior to random subsets, $z = 2.91, p < .01$.

Random subsets of N = 5. Table B-4 presents the results for subsets of $N = 5$. The last line in the table provides multiple regression results based upon the five best predictors (B4 + A4 + A1 + A2 + B3).

Table B-4
Random Subsets of N = 5 vs. the Five Best Predictors

Predictors	Multiple R	Adjusted R^2	$F(5, 710)$	p	SE
A2, A3, A4, B2, B3	.861	.739	406.46	<.0001	99.06
A4, A5, B1, B3, B5	.846	.714	357.40	<.0001	103.81
A2, A3, A4, B1, B5	.854	.728	383.58	<.0001	101.19
A1, A5, B1, B3, B4	.866	.749	426.69	<.0001	97.28
A2, A4, B1, B3, B5	.861	.740	407.67	<.0001	98.95
Mean	.858	.734	396.36		100.06
Best 5	.891	.792	545.02	<.0001	88.51

Five-predictor random subsets accounted, on average, for 73.4% of criterion (TTVIII) variance and produced *SE*s of approximately 100 along with *F* ratios that averaged almost 400. By comparison, the 5 best predictors accounted for 79.2% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple *R* for the five best predictors indicated that the five best predictors were superior to random subsets, $z = 2.57, p < .01$.

Random subsets of N = 6. Table B-5 presents the results for subsets of $N = 6$. The last line in the table provides multiple regression results based upon the six best predictors (B4 + A4 + A1 + A2 + B3 + B2).

Table B-5
Random Subsets of N = 6 vs. the Six Best Predictors

Excluded Predictors	Multiple R	Adjusted R^2	$F(6, 709)$	p	SE
A2, A5, B1, B2	.901	.810	509.11	<.0001	84.55
A2, B3, B4, B5	.892	.795	462.27	<.0001	87.90
A2, A3, B2, B4	.889	.789	447.45	<.0001	89.04
A1, B1, B2, B3	.902	.812	515.63	<.0001	84.12
A1, A2, B1, B5	.897	.803	485.23	<.0001	86.21
Mean	.896	.802	483.94		86.36
Best 6	.915	837	611.18	<.0001	78.42

Six-predictor random subsets accounted, on average, for 80.2% of criterion (TTVIII) variance and produced *SE*s in the 80's, along with *F* ratios that averaged almost 500. By comparison, the 6 best predictors accounted for 83.7% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple *R* for the six best predictors indicated that the six best predictors were superior to random subsets, $z = 2.11$, $p < .05$.

Appendix C
Random Subsets of $N = 2$ Through 9 (AC Data)

Random subsets of $N = 2$. Table C-1 presents the results for subsets of $N = 2$. The first five rows present multiple regression results for the five random subsets. Means in the sixth line of the table are based upon the five random subsets. The cell under the "*p*" column for the "Mean" row is blank because it is meaningless to calculate a mean probability level. The last line in the table provides multiple regression results based upon the two best predictors (A2 + A4).

Table C-1
Random Subsets of $N = 2$ vs. the Two Best Predictors (AC Data)

Predictors	Multiple R	Adjusted R^2	$F(2, 831)$	p	SE
A3, B5	.455	.205	108.59	<.0001	73.23
A1, B2	.506	.254	142.88	<.0001	70.94
A4, B5	.504	.253	141.72	<.0001	71.62
B4, B5	.439	.191	99.41	<.0001	73.88
B1, B3	.406	.162	81.76	<.0001	75.18
Mean	.462	.213	114.87		72.85
Best 2	.630	.395	273.02	<.0001	63.89

Multiple *R's* for the random subsets ranged from .406 to .506, with an average of .462. Two-predictor random subsets accounted, on average, for 21.3% of criterion (TTVIII) variance and produced *SE*s of approximately 72.85 along with significant *F* values generally in excess of 100. By comparison, the two best predictors accounted for 39.5% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple *R* for the two best predictors indicated that the two best predictors were superior to random subsets, $z = 4.92, p < .01$.

Random subsets of $N = 3$. Table C-2 presents the results for subsets of $N = 3$. The last line in the table provides multiple regression results based upon the three best predictors (A2 + A4 + A1).

Table C-2
Random Subsets of $N = 3$ vs. the Three Best Predictors (AC Data)

Predictors	Multiple R	Adjusted R^2	$F(3, 830)$	p	SE
A1, A4, B5	.630	.395	182.23	<.0001	63.90
B2, B3, B5	.507	.254	95.74	<.0001	70.93
A1, B3, B5	.571	.323	133.61	<.0001	67.58
B2, B3, B4	.569	.321	132.27	<.0001	67.69
A3,A4,B4	.639	.406	191.17	<.0001	63.28
Mean	.583	.340	147.00		66.68
Best 3	.724	.523	305.15	<.0001	56.75

Three-predictor random subsets accounted, on average, for 34.0% of criterion (TTVIII) variance and produced *SE*s of approximately 66.68 along with significant *F* values exceeding 100. By comparison, the three best predictors accounted for 52.3% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple *R* for the three best predictors indicated that the three best predictors were superior to random subsets, $z = 5.10, p < .01$.

Random subsets of N = 4. Table C-3 presents the results for subsets of $N = 4$. The last line in the table provides multiple regression results based upon the four best predictors (A2 + A4 + A1 + B4).

Table C-3
Random Subsets of N = 4 vs. the Four Best Predictors (AC Data)

Predictors	Multiple R	Adjusted R^2	$F(4, 829)$	p	SE
A2, A3, A5, B1	.701	.489	200.65	<.0001	58.69
A1, A2, A4, B3	.469	.589	299.97	<.0001	52.63
A5, B2, B4, B5	.613	.372	124.47	<.0001	65.08
A3, A5, B4, B5	.624	.386	131.87	<.0001	64.37
A1, A2, A4, B5	.769	.589	299.76	<.0001	52.64
Mean	.695	.485	211.34		58.68
Best 4	.791	.623	345.61	<.0001	50.41

Four-predictor random subsets accounted, on average, for 48.5% of criterion (TTVIII) variance and produced *SE*s of approximately 59 along with *F* ratios generally in excess of 200. By comparison, the four best predictors accounted for 62.3% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple *R* for the four best predictors indicated that the four best predictors were superior to random subsets, $z = 4.39, p < .01$.

Random subsets of N = 5. Table C-4 presents the results for subsets of $N = 5$. The last line in the table provides multiple regression results based upon the five best predictors (A2 + A4 + A1 + B4 + A3).

Table C-4
Random Subsets of N = 5 vs. the Five Best Predictors (AC Data)

Predictors	Multiple R	Adjusted R^2	$F(5, 828)$	p	SE
A2, A3, A4, B2, B3	.800	.638	295.18	<.0001	49.39
A4, A5, B1, B3, B5	.671	.447	135.45	<.0001	61.11
A2, A3, A4, B1, B5	.777	.601	252.27	<.0001	51.87
A1, A5, B1, B3, B4	.719	.515	177.68	<.0001	57.22
A2, A4, B1, B3, B5	.750	.560	213.72	<.0001	54.47
Mean	.743	.552	214.76		54.81
Best 5	.844	.710	409.25	<.0001	44.22

Five-predictor random subsets accounted, on average, for 55.2% of criterion (TTVIII) variance and produced *SE*s of approximately 55 along with *F* ratios that averaged over 200. By comparison, the five best predictors accounted for 71.0% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple *R* for the five best predictors indicated that the five best predictors were superior to random subsets, $z = 5.70, p < .01$.

Random subsets of N = 6. Table C-5 presents the results for subsets of $N = 6$. The last line in the table provides multiple regression results based upon the six best predictors (A2 + A4 + A1 + B4 + A3 + B2).

Table C-5

Random Subsets of N = 6 vs. the Six Best Predictors (AC Data)

Excluded Predictors	Multiple R	Adjusted R^2	$F(6, 827)$	p	SE
A2, A5, B1, B2	.812	.658	267.67	<.0001	48.06
A2, B3, B4, B5	.805	.645	253.57	<.0001	48.92
A2, A3, B2, B4	.765	.583	194.80	<.0001	53.07
A1, B1, B2, B3	.851	.722	362.07	<.0001	43.29
A1, A2, B1, B5	.786	.615	222.65	<.0001	50.98
Mean	.804	.645	260.15		48.86
Best 6	.892	.793	534.13	<.0001	37.34

Six-predictor random subsets accounted, on average, for 64.5% of criterion (TTVIII) variance and produced *SE*s in the 40's and 50's, along with *F* ratios that averaged over 200. By comparison, the six best predictors accounted for 79.3% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple *R* for the six best predictors indicated that the six best predictors were superior to random subsets, $z = 6.56, p < .01$.

Random subsets of N = 7. Table C-6 presents the results for subsets of $N = 7$. The last line in the table provides multiple regression results based upon the seven best predictors (A2 + A4 + A1 + B4 + A3 + B2 + B5).

Table C-6

Random Subsets of N = 7 vs. the Seven Best Predictors (AC Data)

Excluded Predictors	Multiple R	Adjusted R^2	$F(7, 826)$	p	SE
A5, B4, B5	.893	.795	462.64	<.0001	37.19
B2, B3, B4	.885	.782	427.21	<.0001	38.38
A3, B2, B4	.868	.752	361.50	<.0001	40.92
A2, A5, B2	.843	.708	289.68	<.0001	44.38
A1, A5, B3	.886	.783	430.77	<.0001	38.25
Mean	.875	.764	394.36		39.82
Best 7	.928	.860	730.51	<.0001	30.76

Seven-predictor random subsets accounted, on average, for 76.4% of criterion (TTVIII) variance and produced *SE*s of approximately 40, along with F-ratios that averaged almost 400. By comparison, the seven best predictors accounted for 86.0% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple *R* for the seven best predictors indicated that the seven best predictors were superior to random subsets, $z = 5.99, p < .01$. Thus, unlike the case that obtained with ARNG crews, where random subsets of N = 7 were as effective as the optimal subset of this size, with AC crews randomly selected subsets were less effective than the seven best predictors identified on the basis of stepwise multiple regression procedures.

Random subsets of N = 8. Table C-7 presents the results for subsets of *N* = 8. The last line in the table provides multiple regression results based upon the eight best predictors (A2 + A4 + A1 + B4 + A3 + B2 + B5 + A5).

Table C-7
Random Subsets of N = 8 vs. the Eight Best Predictors (AC Data)

Excluded Predictors	Multiple R	Adjusted R^2	$F(8, 825)$	p	SE
A1, A4	.894	.798	411.32	<.0001	36.96
A1. A3	.901	.810	444.58	<.0001	35.82
B1, B2	.939	.880	766.56	<.0001	28.42
A4, B3	.929	.861	644.91	<.0001	30.65
B2. B5	.930	.863	659.11	<.0001	30.36
Mean	.919	.845	585.29		32.44
Best 8	.960	.920	1201.32	<.0001	23.21

Eight-predictor random subsets accounted, on average, for 84.5% of criterion (TTVIII) variance and produced *SE*s of approximately 32, along with F-ratios that averaged almost 600. By comparison, the eight best predictors accounted for 92.0% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple *R* for the eight best predictors indicated that the eight best predictors were superior to random subsets, $z = 7.54, p < .01$. Thus, unlike the case that obtained with ARNG crews, where random subsets of N = 8 were as effective as the optimal subset of the same size, with AC crews randomly selected subsets were less effective than the eight best predictors identified on the basis of stepwise multiple regression procedures.

Random subsets of N = 9. Table C-8 presents the results for subsets of *N* = 9. The last line in the table provides multiple regression results based upon the nine best predictors (A2 + A4 + A1 + B4 + A3 + B2 + B5 + A5 + B3).

Nine-predictor random subsets accounted, on average, for 93.8% of criterion (TTVIII) variance and produced *SE*s of approximately 20, along with F-ratios that averaged approximately 1,566. By comparison, the eight best predictors accounted for 96.5% of criterion variance. A test between the mean Multiple *R* for the random subsets and the Multiple *R* for the nine best predictors indicated that the nine best predictors were superior to random subsets, $z = 5.93, p < .01$. Thus, unlike the case that obtained with

ARNG crews, where random subsets of N = 9 were as effective as the optimal subset of the same size, with AC crews randomly selected subsets were less effective than the nine best predictors identified on the basis of stepwise multiple regression procedures.

Table C-8

Random Subsets of N = 9 vs. the Nine Best Predictors (AC Data)

Excluded Predictor	Multiple R	Adjusted R^2	$F(9, 824)$	p	SE
B5	.973	.946	1611.70	<.0001	19.15
A4	.958	.916	1015.80	<.0001	23.75
A5	.973	.946	1627.74	<.0001	19.06
B4	.958	.917	1020.66	<.0001	23.70
B1	.983	.965	2555.44	<.0001	15.36
Mean	.969	.938	1566.27		20.20
Best 9	.983	.965	2555.44	<.0001	15.36

Thus, regardless of subset size, more predictive power was achieved by following the engagement selection strategy that was revealed by stepwise multiple regression procedures. The discrepancy in predictive power between random and optimal subsets, however, progressively diminishes as more engagements are added to the prediction equation. With nine-engagement subsets, for example, the difference in predictive power between random and optimal subsets is statistically significant, but of little practical significance. Nonetheless, for subsets of any size using this set of AC data, best results were obtained by using regression-determined combinations of engagements.